VIDEO SUPER-RESOLUTION BY INTEGRATING SAD AND NCC MATCHING CRITERION FOR MULTIPLE MOVING OBJECTS

Chen-Chiung Hsieh*, Yo-Ping Huang*, Yu-Yi Chen*, Chiou-Shann Fuh**, and Wen-Jen Ho***
** Department of Computer Science and Engineering, Tatung University, Taipei, Taiwan R.O.C.*
*** Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan R.O.C.*
**** Institute for Information Industry, R.O.C.*
cchsieh@ttu.edu.tw

ABSTRACT
Traditionally, image enlargement is magnified from a single image. Due to the one and only one image, the quality of the reconstructed image is thus constrained. Super-resolution is proposed to use multiple frames as additional information to estimate the high-resolution image. If we have enough low-resolution images with observed subpixels, the high-resolution image can be reconstructed. In this paper, a complete super-resolution model based on k-means motion clustering is proposed for image enlargement with multiple moving objects. Motion masks are first produced for useful image selection and then blocks matching are used to do motion estimation. Two complementary features, sum of absolute difference (SAD) and normalized cross-correlation (NCC), are adopted as the matching criterion. Objects are assumed to move slightly between two consecutive images. Thus, erroneous motion vectors could be corrected by the center of motion clusters. The proposed method achieves better magnification quality than the traditional ones and some previous super resolution works. From the experimental results, both the visual and quantitative improvements, especially in the high frequency, are significant.

KEY WORDS
Super-resolution, image enlargement, motion estimation, block matching, k-means clustering

1. Introduction

A video camera has limited spatial resolution which depends on the number of charge-coupled device (CCD) image sensors. Due to the higher cost with hardware technology, software approach is usually adopted to enlarge image. Interpolation-based approaches are the basic techniques in image enlargement and they involve selection of data values from the known pixels such as the nearest-neighbor interpolation, bilinear interpolation, and bi-cubic interpolation. However, these methods have constraint on the number of available pixels in a single image. For the sake of additional information, multiple images are considered to resolve such problems. The basic idea is illustrated in Figure 1.

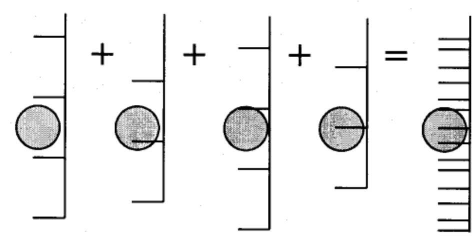


Figure 1: The basic idea of super-resolution.

Multiple frames based approach is to reconstruct a single high-resolution image as shown in the right part of Fig. 1 from a sequence of low resolution video images as shown in the left part of Fig. 1, and is sometimes referred to as super-resolution (SR). It is a useful method for generating a high-resolution (HR) image from multiple overlapping low-resolution (LR) images of translated scene [1, 2, 3, 4, 5, 6, 7, 8, 9]. It works only if the frames are shifted by fractions of a pixel from each other. The super-resolution algorithm is to aggregate those registered sub-pixels contained in the smaller original frames to produce a larger image.

However, some restrictions itemized as followings are made with regarding to different applications.

- **Movement:** If the objects in image are completely still or only move a little, the super-resolution image could not be well reconstructed for the lack of exposed sub-pixels.
- **Sampling rate:** On the other hand, the captured images may be blurred and not useful to super-resolution when the objects are moving too fast [10]. Therefore, the temporal sampling rate could not be too slow.
- **Illumination:** When the original images are captured with much noise, overexposure, or underexposure, we could not reconstruct the good high-resolution images for the error-prone image registration.

High-resolution images are useful and important in many applications. For example, the high-resolution image could reveal more detail and important information in medical applications, astronomical imaging, and video surveillance [10]. In recent years, videos from cheap low quality cameras increase more and more, and the demand of higher resolution from these low resolution images attracted a lot of attentions.

2. Related Works

The overall system for enhancing image resolution of video is mainly composed of two components. The first and the most important step is motion estimation which takes a set of low-resolution frames as input and produces motion vectors from one to another frame. The second is the super-resolution algorithm which combines low-resolution frames and motion vectors to reconstruct a high resolution image. Some existing works involved with the system are briefly described in the following.

2.1 Non-uniform Interpolation Approach

This approach [11] is the most intuitive method for SR image reconstruction. There are three stages to be performed successively as below:

1. Motion Estimation: i.e., registration (if the motion information is not known).
2. Non-uniform Interpolation: adopted to produce an improved resolution image.
3. De-blurring process: dependent on the observation model.

With the estimated motion information, the HR image on non-uniformly spaced sampling points could then be obtained. The direct or iterative reconstruction procedure is followed to produce uniformly spaced sampling points [12, 13, 14, 15].

2.2 Bayesian MAP Method

Schultz and Stevenson [16] introduced the Bayesian maximum a posteriori probability (MAP) method for image enlargement. The reconstruction of the super-resolution image is placed into a statistical framework by using the MAP estimation. Since super-resolution is an ill-posed problem, this method incurring a-priori constraints to transform the problem into a well-posed problem. Experimental results in [16] showed that the performance of MAP method is superior to that of the bilinear and bicubic interpolation methods.

2.3 Iterated Back Projection (IBP) Method

Irani and Peleg [2] introduced the iterated back projection method. The high-resolution image reconstruction is accomplished by iteratively minimizing the difference between the given low-resolution images and the simulated low-resolution images generated from down-sampling the current guessed high-resolution image. This kind of method is adopted in our proposed system.

3. System Architecture

The proposed system is mainly composed of the initial guess, motion estimation and the reconstructing high-resolution image as shown in Figure 2. In the following, we will discuss how these functional blocks reconstruct the high-resolution images.

We have to remove noise in the beginning because super-resolution will also enhance noise. Usually, median filter or Wiener filter are adopted to smooth out noise.

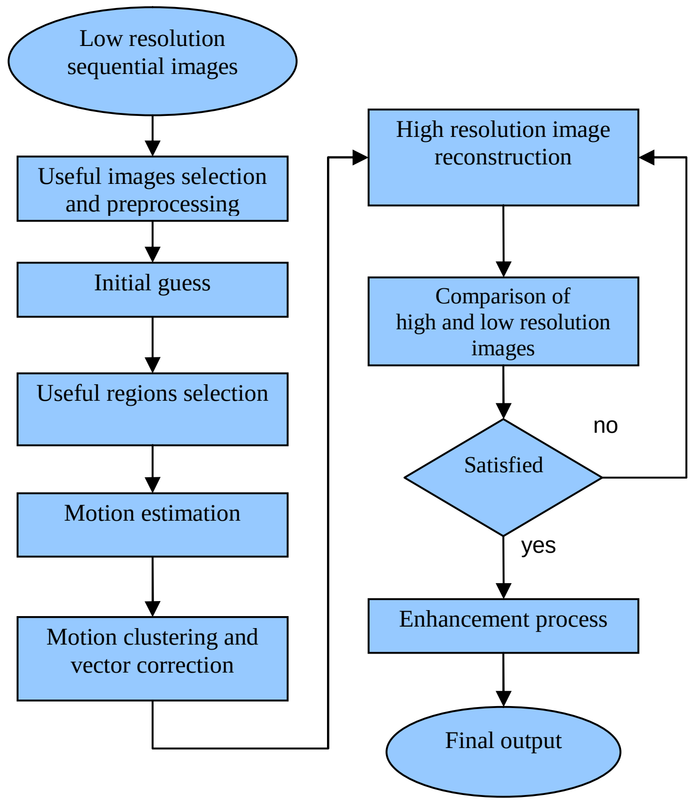


Figure 2: The flowchart for the proposed system.

3.1 Initial Guess

After preprocessing, we choose an interesting frame to do super-resolution. This frame is so called the reference frame and it can be magnified initially by the bi-cubic interpolation.

When the magnification factor and the reconstruction image size become larger, the computation time gets longer. The initial guess [3] will largely affect the performance of our system. If we have good initial guesses, the process of super-resolution will be faster. On the contrary, the result of super-resolution will be worse if the initial guess is also worse.

3.2 Motion estimation

Many low-resolution images are taken from digital camera. However, some of these images are duplicate or even without motions. Those useless images will increase the runtime of our system. For this reason, we can remove them by comparing the *n* th image and the *n+1* th image in the low-resolution image sequence.

The difference between the current image and the reference image, and the difference between the current image and the previous image are computed. Then, the two difference images are intersected to produce moving regions. Non-moving region and the occlusion boundary are excluded to reduce the executing time. Figure 3 shows the useful moving regions

Motion estimation [17, 18] plays a very important role in super-resolution. According to the magnification factor, we need to shift reference image to different scale of subpixels. Each shifted image represents a different subpixel moving direction. These images of the different subpixel movements are compared with the current image. We segment image into 8*8 blocks and motion estimation is then done by block matching.

The matching criterion is the sum of absolute difference (SAD) and normalized cross-correlation (NCC). The SAD, between a block in the current frame *k* and the block after it displaced by a motion vector $d=(u,v)$ in the reference frame *l*, is formulated as in Eq. (1).

$$SAD = \sum_{x,y \in B} \left| I_k(x,y) - I_l(x+u, y+v) \right|, \qquad (1)$$

where $I(x,y)$ denotes the pixel value at (x,y) within the matching block *B*. The estimated motion vector is then given as $d=(u,v)|SAD_{\min}$ for all possible locations within the search window.

Another criterion is the normalized cross-correlation (NCC) which resists brightness better. The correlation between two signals (cross-correlation) is a standard approach for feature detection. The NCC stresses the features of image and is formulated as in Eq. (2).

$$r(u,v) = \frac{\sum_{x,y \in B} [f(x,y) - \bar{f}_{u,v}][t(x+u,y+v) - \bar{t}\,]}{\left\{ \sum_{x,y \in B} [f(x,y) - \bar{f}_{u,v}]^2 \sum_{x,y \in B} [t(x+u,y+v) - \bar{t}\,]^2 \right\}^{0.5}}, \qquad (2)$$

where *t* is the template and *f* is the search region. $\bar{t}$ is the mean of the template and $\bar{f}_{u,v}$ is the mean of $f(x,y)$ in the search region under the template, and $r(u,v)$ is the correlation coefficient at point (u,v) for *f* and *t*. The estimated motion vector $d=(u,v)$ is given as $d=(u,v)|NCC_{max}$.

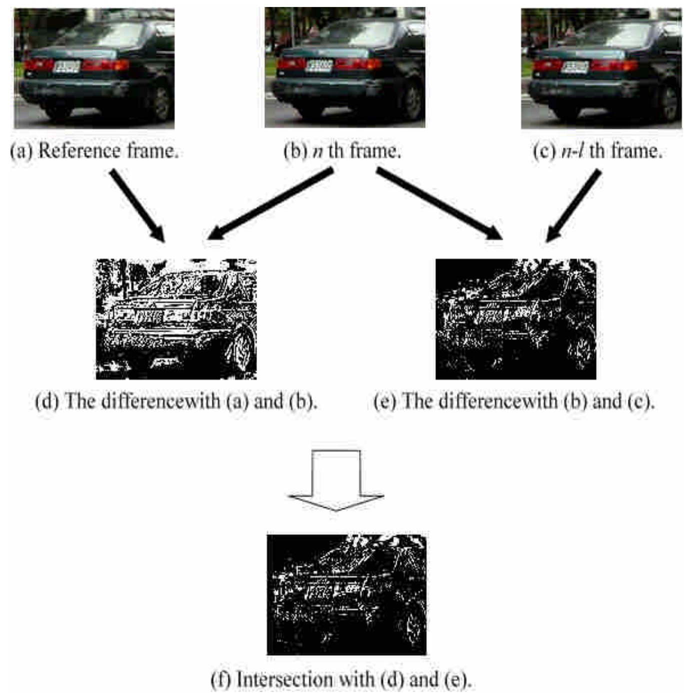


Figure 3: Selection of useful moving regions.

In this paper, only blocks exceeded the given threshold would be record. However, super-resolution quality would be decreased if the motion vectors are estimated wrong. Thus, we need to do motion clustering for motion correction.

3.3 Motion Clustering and Motion Correction

After motion estimation, the motion vectors are clustered for each individual object. The k-means algorithm is adopted to assign each motion vector to the cluster whose center is nearest. The center is the average of all the motion vectors belonging to the cluster. Each cluster consists of many similar motion vectors. By assuming rigid objects and moving slightly, the mismatched motion vectors within clusters can therefore be removed.

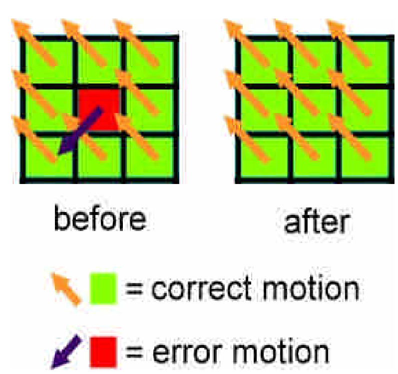


Figure 4: Motion vector correction.

After removing the mistakes of motion estimation by clustering, there are still few mistakes in the motion vectors for each frame. Therefore, the motion vector of each block is examined. If the maximum number of the same motion vectors around the current block exceeds a given threshold, the motion vector of the current block could be corrected to the maximal one. Figure 4 shows the correction step.

3.4 High Resolution Image Reconstruction

Super-resolution reconstructs the high-resolution image from the low-resolution images. We can use the corrected motion vectors of each low-resolution image to do super-resolution. By comparing the real low-resolution images with the simulated low-resolution images that are produced by the *n* th super-resolution image. Then, the differences are used to improve the *n+1*th super-resolution image. We repeatedly apply the super-resolution process until the reference frame converges to a satisfactory result after several iterations. The imaging process of g_k at *n* th iteration is simulated by Eq (3).

$$g_k^{(n)} = (T_k(f^{(n)}) * h) \downarrow s, \qquad (3)$$

where g_k is the k th observed image frame; f is the super-resolution image; h is the blurring operator defined by point-spread-function (PSF) of the sensor of the digital camera; T_k is the transformation operator that transforms other low-resolution images to the reference frame; and $\downarrow s$ is the down-sampling operator. The iterative scheme of the high-resolution is updated by Eq (4).

$$f^{(n+1)} = f^{(n)} + \frac{1}{K}\sum_{k=1}^{K} T_k^{-1}(((g_k - g_k{}^{(n)}) \uparrow s) * p) \ ,(4)$$

where K is the total number of low-resolution images that are used; p is the de-blurring operator; $\uparrow s$ is the up-sampling operator; and $f^{(n)}$ is the reconstruction result after *n* th iterations.

After super-resolution, the reconstructed high-resolution image may have noise or be dark. In order to make it clearer and more recognizable, the post-processing is recommended. To recover the sharpness that was smoothed by preprocess, sharpening the edge is necessary. It can greatly enhance the part of detail. Besides, adaptive histogram equalization is used to make the image easier to view.

If the high-resolution image is not clear, those image enhancement techniques are helpful for human recognition. Finally, we can output the refined high-resolution image.

4. Experimental Results and Discussions

Experimental results of the proposed system are demonstrated to show its feasibility for real applications. The experimental environments are described as follows:

- Personal computer (PC): Intel 3.0 GHz CPU and 1024 MB DDR RAM
- Capture device: digital video camera with low resolution ranging from 100*100 ~ 300*300.
- Magnification factor: 2 times and then 2 times.

4.1 Reconstruct HR Images with Simulated Images

By taking an image as the original scene, move it in sub-pixel and down-sample it into several images. Then, we take the simulated images as input and reconstruct the high-resolution image with magnification factor of 4. The result is shown in Figure 5.

We use the PSNR to evaluate the result. The PSNR is formulated as in Eq (5).

$$\begin{aligned} &\text{MSE} = \sum \frac{[f(i,j) - F(i,j)]^2}{M * N}, \\ &\text{RMSE} = \sqrt{\text{MSE}} \\ &\text{PSNR} = 20\log_{10}\left(\frac{255}{\text{RMSE}}\right) \end{aligned} \qquad (5)$$

In Eq (5), the $f(i,j)$ is the pixel value of the point (i,j) at f image and the $F(i,j)$ is the pixel value of the point (i,j) at F image. f is the original image and F is the processed image. The quantitative results are shown in Table 1. We can discover that the performance is good and the super-resolution image is better than the others.

Table 1: The PSNRs for techniques with simulated image.

Technique	PSNR
The zero-order interpolation	26.5679
The bilinear interpolation	27.0679
The bilinear interpolation	27.4232
The proposed method	27.6182

4.2 Reconstruct HR Images with Multiple Moving Objects.

We take 30 pictures of the real multiple moving objects, moving cars on road by a camera. There were three objects in the video: the blue one in front, the big one moved fast behind, and a smaller motorist on occlusion boundary. The result is shown in Figure 6 with the multiple moving objects and magnification factor 2.

4.3 Reconstruct HR Images with Moving Camera

The video shows that there is a red car still in road side. The image sequence was down-sampled to do super-resolution. We use high pass filter to get the region of the high frequency. Results of the experiments are shown in Figure 7 and evaluated by measuring PSNR. Table 2 shows the difference of the high frequency produced by different technique with the original HR image.

Table 2: The PSNR for each technique in high frequency region.

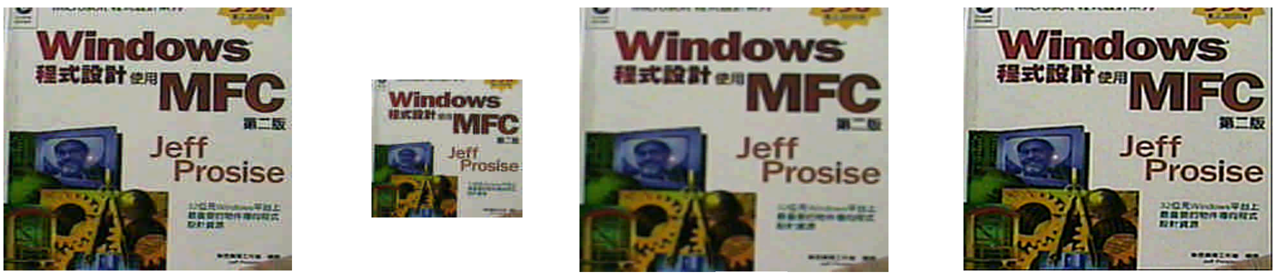
Technique	PSNR	Difference
The zero-order interpolation	16.8937	22.77%
The bilinear interpolation	17.4835	19.54%
The bicubic interpolation	17.8292	19.09%
The super-resolution method	18.5311	17.91%

4.4 Evaluation and Analysis

For reconstructing high-resolution images with simulated moving image, the reconstructed HR image computed from the observed low-resolution sequence showed great improvements over the traditional zero-order, bilinear, and bicubic interpolation. The quantitative results also show that the proposed method produce higher PSNR values than the others.

For the multiple moving objects, the reconstructed HR image is good, but the rear white car moved too fast to cause motion blur in the result. And the other car is near occlusion boundary, so it was not reconstructed by the method.

The proposed super-resolution image is greatly enhanced visually with moving video camera. And the PSNR value of the super-resolution image is better than the others, ranging from 0.7 to 1.6 dB. The difference values of the high frequency region are also better from 1% to 5%. From the experimental results, the quality of the images, especially the edges, has a great improvement by the proposed super-resolution algorithm.



(a) (b) (c) (d)

Figure 5: Reconstructed high resolution image with the simulated images. (a) The original high-resolution image. (b) One of the simulated low-resolution images. (c) High resolution image by the bicubic interpolation method. (d) High resolution image by the proposed super-resolution method.

(a) (b) (c) (d)

Figure 6: Reconstructed high resolution image from real LR images with multiple moving objects. (a) Image expanded by the zero-order interpolation method. (b) Image expanded by the bilinear interpolation method. (c) Image expanded by the bicubic interpolation method. (d) Image expanded by the proposed super-resolution method.

(a) (b) (c)

Figure 7: Comparison the high frequency region among different approaches. (a) The high frequency region of the original high-resolution image. (b) The high frequency region of the bicubic interpolation. (c) The high frequency region of the super resolution.

5. Conclusions

We purposed a super-resolution algorithm which can deal with the multiple moving objects with SAD and NCC used as matching criterion. Meanwhile, motion vectors could be corrected by k-means motion clustering and a reference image was chosen to do super-resolution. If we were interested in some region, we can only reconstruct the specific region. When we determine the reference frame, we will de-noise all the captured frames and remove the duplicate ones. The super-resolution consists of three major steps: initial guess, motion estimation, and reconstructing the high-resolution image with iterative method.

First, the initial high-resolution image was guessed by the bicubic interpolation. We could shift reference image to different scale of subpixels that are needed. Secondly, the current image was segmented into blocks for motion estimation. By using both SAD and NCC, complements each other, the block matching was good with respect to the brightness, color, and value. The non-moving regions were removed by the difference frame among the reference frame, the *n-1* th frame, and the *n* th frame to reduce the executing time. K-means motion vector clustering was used to extract clusters of moving object and then the error vectors could be corrected. Finally, the high-resolution image was reconstructed by iterative method. Repeatedly apply the super-resolution process until the reference frame converged to a satisfactory result. Image enhancement was adopted to improve the quality of the SR image.

The proposed method is better than the traditional ones and could deal with image sequence of multiple moving objects. From the experimental results, our method is very useful, especially in terms of high frequency parts, for image sequence enlargement. The PSNR of the proposed method is better than the others, ranging from 0.7 to 1.6 dB. The difference values of the high frequency regions are also better ranging from 1% to 5%. To save the executing time for processing the whole scene, motion regions can be used to speed up in motion estimation. In other words, block matching candidates can be greatly reduced. However, objects undergoing unrestricted motion can be more complex and motion vectors clustering needs to be refined in the future

References

[1] M. Irani and S. Peleg, Improving resolution by image registration, *CVGIP: Graphical Models and Image Processing, 53*, 1991, 231-239.

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

[3] C.Y. Chen, Y.C. Kuo, and C.S. Fuh, Image reconstruction with improved super-resolution algorithm, *International Journal of Pattern Recognition and Artificial Intelligence, 18*(8), 2004, 1-15.

[4] S. Chaudhuri and D.R. Taur, High-resolution slow-motion sequencing: how to generate a slow-motion sequence from a bit stream, *IEEE Signal Processing Magazine, 22*(2), 2005, 16 - 24.

[5] M. Elad and Y. Hel-Or, A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur, *IEEE Transactions on Image Processing, 10*(8), 2001, 1187- 1193.

[6] B.K. Gunturk, A.U. Batur, Y. Altunbasak, M.H. III Hayes, R.M. Mersereau, Eigenface-domain super-resolution for face recognition, *IEEE Transactions on Image Processing, 12*(5), 2003, 597-606.

[7] H. Shen, P. Li, L. Zhang, and Y. Zhao, A MAP algorithm to super-resolution image reconstruction, *Proc. 3rd International Conference on Image and Graphics*, Hong Kong, China, Dec. 2004, 544–547.

[8] B.K. Gunturk, Y. Altunbasak, and R.M. Mersereau, Multiframe resolution- enhancement methods for compressed video, *IEEE Signal Processing Letter, 9*, June 2002, 170-174.

[9] A.J. Patti, M.I. Sezan, and A.M. Tekalp, Super resolution video reconstruction with arbitrary sampling lattices and nonzero aperture time, *IEEE Trans. Image Processing, 6*(8), 1997, 1064-1076.

[10] M. Ben-Ezra, A. Zomet, and S.K. Nayar, Video super-resolution using controlled subpixel detector shifts, *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*(6), June 2005, 977-987.

[11] S.C. Park, M.K. Park, and M.G. Kang, Super-resolution image reconstruction: a technical overview, *IEEE Signal Processing Magazine, 20*(3), May 2003, 21-36.

[12] J.J. Clark, M.R. Palmer, and P.D. Laurence, A transformation method for the reconstruction of functions from nonuniformly spaced samples, *IEEE Trans. Acoustics, Speech and Signal Processing, 33*(5), 1985, 1151-1165.

[13] S.P. Kim and N.K. Bose, Reconstruction of 2-D bandlimited discrete signals from nonuniform samples, *IEE Proc. Radar Signal Processing, 137*, June 1990, 197-204.

[14] A. Papoulis, Generalized sampling theorem, *IEEE Trans. Circuits Systems, 24*(11), Nov. 1977, 652-654.

[15] J.L. Brown, Multi-channel sampling of low pass signals, *IEEE Trans. Circuits Systems, 28*, Feb. 1981, 101-106.

[16] R.R. Schultz and R.L. Stevenson, Extraction of high-resolution frame from video sequence, *IEEE Transactions on Image Processing, 5*(6), 1996, 996-1011.

[17] V. Argyriou and T. Vlachos, Sub-pixel motion estimation using gradient cross-correlation, *Proc. 7th International Symposium on Signal Processing and its Applications*, Paris, France, July 2003, 215-218.

[18] J.-W. Suh and J. Jeong, Fast sub-pixel motion estimation techniques having lower computational complexity, *IEEE Transactions on Consumer Electronics, 50*(3), Aug. 2004, 968-973.