**SUPER RESOLUTION: A SPATIAL SUBPIXEL INTERPOLATION TECHNIQUE**

Eng. Popescu Andrei-Iulian,

Computer Science and Engineering Department, Faculty of Automatic Control and Computers,

Politehnica University of Bucharest, Splaiul Independence 313, Bucharest 060042, Romania,

[andrei.popescu96@stud.energ.upb.ro](mailto:andrei.popescu96@stud.energ.upb.ro)

Eng. Busuioc Dragos-Andrei,

Computer Science and Engineering Department, Faculty of Automatic Control and Computers,

Politehnica University of Bucharest, Splaiul Independence 313, Bucharest 060042, Romania,

[dragos.busuioc@stud.acs.upb.ro](mailto:dragos.busuioc@stud.acs.upb.ro)

Professor PhD Eng. Costin Anton Boiangiu,

Computer Science and Engineering Department, Faculty of Automatic Control and Computers,

Politehnica University of Bucharest, Splaiul Independence 313, Bucharest 060042, Romania,

[costin.boiangiu@cs.pub.ro](mailto:costin.boiangiu@cs.pub.ro)

**Abstract**

Super resolution describes the process of creating a higher resolution image using one or multiple low resolution images. In this paper we will focus on the method based on multiple low resolution images, presenting its advantages and disadvantages. Finally, we will present the challenges which future research is faced with.

The need for this field stems from applicability in important areas such as machine image perception and human image interpretation, where high end image capturing hardware isn’t feasible or cost efficient.

**Introduction**

**Problem motivation**

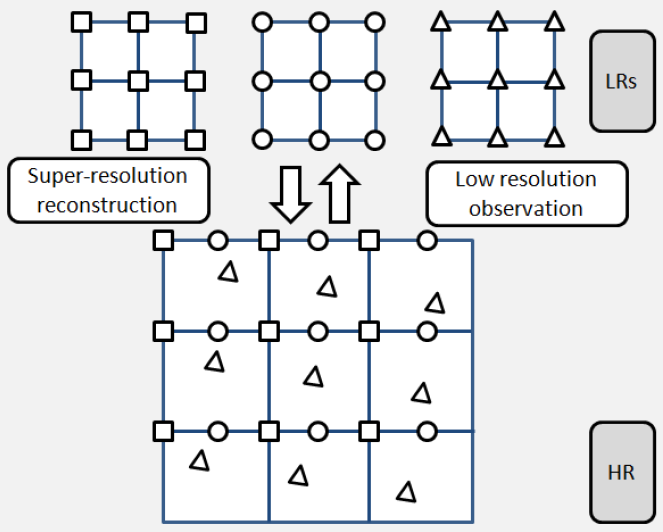
The main limitation of image acquisition is the imaging sensors of the device used. The modern image sensor is typically a charge-coupled device (CCD) or a complementary metal-oxide-semiconductor (CMOS) active-pixel sensor. The two types are typically arranged in a two-dimensional array to capture two-dimensional image signals. The spatial resolution of the image to capture is directly influenced by the sensor size or equivalently the number of sensor elements per unit area. Higher density of the sensors means higher spatial resolution possible for the imaging system. An imaging system with inadequate detectors will generate low resolution images with blocky effects, due to the aliasing from low spatial sampling frequency. To increase the sensor density will bring issues such as causing “shot noise” [1], increasing hardware cost and sensor size.

While the image sensors limit the spatial resolution of the image, the image details (high frequency bands) are also limited by the optics, due to lens blurs (associated with the sensor point spread function (PSF)), lens aberration effects, aperture diffractions and optical blurring due to motion. The construction of imaging chips and optical components able to capture very high-resolution images is expensive and not practical in most real applications. One example is the widely used surveillance cameras. Besides the cost, the resolution of a surveillance camera is limited in the camera speed and hardware storage.

Another way to address this problem is to accept the image degradations and use signal processing to post process the captured images, to trade off computational cost with the hardware cost. These techniques are specifically referred as Super-Resolution (SR) reconstruction.

Super-resolution (SR) us the concept of constructing high-resolution (HR) images from several observed low-resolution (LR) images, thus increasing the high frequency components and removing the degradations caused by the imaging process of the low resolution camera. By combining the non-redundant information contained in multiple low-resolution frames, we can generate a high-resolution image. A closely related technique with SR is the single image interpolation approach, which can be used to increase the image size, but without additional information provided, the quality of the single image interpolation is limited due to the ill-posed nature of the problem and in some cases add no value. In the SR setting, however, multiple low-resolution observations are available for reconstruction.

The non-redundant information contained in the these LR images is typically introduced by subpixel shifts between them. These shifts may occur due to movements of objects, or due to controlled motions, like the imaging system having a predefined speed and path. Each low-resolution frame is a decimated, aliased observation of the true scene, thus SR is possible only if subpixel motions between these low resolution frames exist.



[2]

In the imaging process, the camera captures several LR frames, which are down sampled from the HR scene with subpixel shifts between each other. SR construction reverses this process by aligning the LR observations to subpixel accuracy and combining them into a HR image grid(interpolation), thereby overcoming the imaging limitation of the camera. SR (some of which described in this book), arises in many areas such as:

1. Surveillance video [3][4]: frame freeze and zoom on regions of interest (usages consist of automatic target recognition and human perception: license plates, faces etc.)
2. Medical imaging (CT, MRI, Ultrasound) [5][6]

The image super-resolution (SR) idea was introduced by Tsai and Huang [7] to reconstruct the HR image by recovering additional information from multiple LR images using signal processing techniques. The SR approach has attracted increasing research attentions in recent years, and many improved methods such as single image SR and real-time SR, have been reported [8,9].

**Related work**

Early works regarding multiframe super resolution are based upon classical flow estimation techniques with differential equations (Lucas & Kanade. 1981 [10] on coarse-tofine pyramidal approximation and Hom & Schunck, 1981 [11] on regularized improvement) in order to achieve better low-resolution images with improved subpixel precision.

Frequency-based SR. Frequency-based SR methods were first proposed in the mid-1980s [12]. Tsai and Huang tried to de-alias LRIs by exploiting phase differences among them. This first algorithm operated on observations that are free from degradation and noise. Though computationally attractive, frequency-based methods have significant disadvantages as the assumption of ideal sampling is unrealistic and observation noise is not considered. Most importantly, their global translation model is, for many applications, inappropriate.

The frequency-domain modeling of rigid transformation has been tackled by Vandewalle et al. (2016) [13], but results have proven that these approaches are not suitable for video sequences with complex nonrigid motion. However, recent advances in optical flow estimation manage to provide subpixel accuracy even when low-resolution frames were captured on a complex motion, as shown in D. Sun, Roth, and Black (2010) [14] and Baker et al. [15]. This was achieved by exploiting nonlocal regularization as a generalization of median filtering for outlier rejection on a model stemming from Horn and Schunck (1981). Contemporary optical flow methods can improve accuracy with respect to classical approaches, as the motion becomes more complex, contemporary methods clearly outperform the classical ones, thanks to a greatly reduced number of outliers.

Recent methods allow a greater reduction of outliers, thanks to the edge preserving fusion of sparse feature-based matches with continuous flow fields as in the recent work by Revaud, Weinzaepfel, Harchaoui, and Schmid (2015).

The simpler and faster block matching techniques do not suffice to accurately describe the subpixel motion between frames, although there are ways to improve accuracy based on iterative refinement of matches which can provide reasonable approximation of optical flow fields.

(Salvador, Kochale, & Schweidler, 2013).

Learning-based SR algorithms have gained significant popularity since their introduction around 2000 [16]. They incorporate strong application-dependent priors in order to reconstruct HRI. One of the most frequently used approaches is so-called ”hallucination” reconstruction. This is the process of obtaining a high-resolution image (usually of a human face) from an input low-resolution image (or images, although an SR result can be obtained even from a single image), with the help of a large collection of other high-resolution images (a training set).

**Current State-of-the Art**

Recently, several models based on deep neural networks have achieved great success in terms of both reconstruction accuracy and computational performance in regards to super-resolution. [17,18]

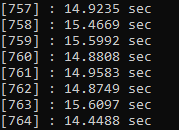
EDSR[19] - In particular, residual learning techniques exhibit improved performance. Recent models of EDSR have proven significant performance by removing unnecessary modules in conventional residual networks and expanding the model size while stabilizing the training procedure.

Fast Super-Resolution Convolutional Neural Network (FSRCNN)[20,21] - This method is an improved Super-Resolution Convolutional Neural Network (SRCNN) that manages to achieve real-time super-resolution while still maintaining good performance.

Deep Laplacian Pyramid Super-Resolution Network (LapSRN)[22,23] - Provides fast and accurate image super-resolution, the proposed network progressively reconstructs the sub-band residuals of high-resolution images at multiple pyramid levels.

**Proposed method**

The current project is expected to bring contributions to the computational speed of individual frames or an improvement in the overall quality of the concerned image sequence. The process will take in consideration multiple optical flow methods as well as varying input parameters. The speed improvement will be focused on using Nvidia CUDA processes to achieve parallelism, it will not only compute frames faster, but it will also open the possibility of real-time super resolution.

**Performance Measurements**

**Preliminary Results**

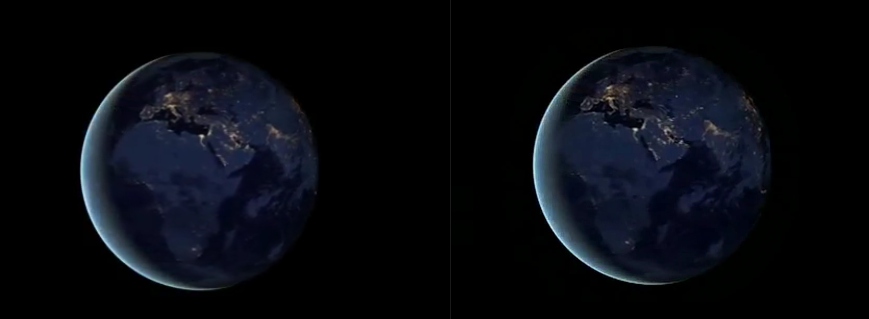
As a preliminary result we used the following input parameters:

* Input Video: planet.avi [480 x 270]
* Scale factor: x2
* Iterations:10
* Temporal radius: 4
* Optical Flow: Farneback
* Mode: CPU

During runtime we encounter considerable process time for individual frames: ~ 15sec /frame . Considering the average video is filmed at 30FPS, it will take one hour to create a supper resolution version of an 10 seconds video.

**Preliminary Conclusions**

To show the super resolution, the first frame of the planet.avi and planetResult.avi videos are compared with and without super resolution.



In the right-hand-side section of the preceding figure, you can observe more details in the aspect of the continents and lights as the video was upscaled.

However, the downside remains the processing time in which the supper resolution is achieved.

**Demonstrator application architecture details**

Our current application takes a video source through the following processing flow:

1. Reads the input parameters
   1. Input video path
   2. Output video path
   3. Scale factor
   4. Iteration count
   5. Radius of temporal search area
   6. Optical flow algorithm
      1. Farneback[24]
      2. TVL1[25]
      3. Brox[26]
      4. PRYLK[27]
   7. Type of processing unit
2. Applies the selected algorithm
   1. Farneback
   2. TVL1
   3. Brox
   4. PRYLK
3. Saves the output at the designated path

In the **Farneback** algorithm, the local signal model is represented in a local coordinate system by the function

,

where A is a symmetric matrix, b a vector and c a scalar. The application determines, based on their position in the neighborhood, the relative weight of points.

Based on this function, we calculate what happens if the image were to be displaced, the function becoming

,

Where d is the displacement, which is

The **TV-L1** algorithm is based on the functional equation

The generally non-convex energy functional becomes a convex minimization problem after linearization of the image intensities, but this linearization is only valid for small displacements. Image pyramids with a downsampling factor of 2 are employed. Beginning with the coarsest level, the equation is solved at each level of the pyramid and the solution is propagated to the next ﬁner level.

In the **Brox** algorithm, a variation model is used to solve the optical flow method. It goes with the following constraints as a premise:

* Grey value constancy assumption
* Gradient constancy assumption
* Smoothness assumption
* Multiscale approach

For the value calculations, a Euler-Lagrange equation is used

, which is approximated numerically with the use of fixed point iterations on

**Preliminary performance measurements**

After running the algorithms on a few input videos (considering a 640x368 resolution video, upscaling x2 and 10 iterations per frame) the average computational time per frame was:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Farnback** | **TV-L1** | **Brox** | **PRYLK** |
| **CPU** | **15.48s** | **53.785s** | **27.1s** | **28.376s** |
| **GPU** | **N/A** | **3.482s** | **0.7217s** | **N/A** |

**Final performance review**

**Conclusions**

**Future work**

A next step in the solution development could be found by taking one of the following approaches:

* post processing enhancements. In latest years, Nvidia has been using Dynamic Super Resolution in their solutions, using such a technology to reduce the blur caused by upscaling could prove useful.
* Experimenting with new algorithms

**References**

[1] Mohd Awais Farooque, Jayant S.Rohankar. SURVEY ON VARIOUS NOISES AND TECHNIQUES FOR DENOISING THE COLOR IMAGE, pages 1-2, 2013

[2] Jianchao Yang, Thomas Huang, Image super-resolution: Historical overview and future challenges, pages 6, 2017

[3] Marco Crisani, Dong Seon Cheng, Vittorio Murino, and Donato Pannullo. Distilling information with super-resolution for video surveillance. In Proceedings of the ACM 2nd International Workshop on Video Surveillance and Sensor Networks, pages 2-11, 2004

[4] Frank Lin, Clinton B. Fookes, Vinod Chandran, and Sridha Sridharan. Investigation into optical flow super-resolution for surveillance applications. In The Austrilian Pattern Recognition Society Worshop on Digital Image Computing, pages 987-995, 2005

[5] J. Maintz and M. Viergever. A survey of medical image registration. Medical Image Analysis, pages 1-36, 2008

[6] S. Peleg and Y. Yeshurun. Superresolution in MRI: application to human white matter fiber tract visualization by diffusion tensor imaging. Magnazine Resonance in Medicine, pages 29-35, 2001.

[7] S. Baker and T. Kanade. Hallucinating faces. In Fourth International Conference on Automatic Face and Gesture Recognition, pages 83–88, 2000.

[8] R. Tsai and T. Huang, ‘‘Multiframe image restoration and registration, Advances in Computer Vision and Image Processing (JAI Press), pp. 317–339, 1984.

[9] S. Borman and R. Stevenson, ‘‘Super-resolution from image sequences-a review,’’ Proc. of Midwest Symposium on Circuits and Systems, pp. 374–378, 1998.

[10] Lucas, B. D., & Kanade, T. (1981). An iterative image registration technique with an application

to stereo vision. In Proceedings of the international joint conference on artificial intelligence

(p. 674–679).

[11] Horn, B. K. P., & Schunck, B. G. (1981). Determining optical flow. Artificial Intelligence, 17, 185–203

[12] R. Y. Tsai and T. S. Huang. Multiframe image restoration and registration. Advances in Computer Vision and Image Processing, 1:317–339, 1984.

[13] Vandewalle, P., Süsstrunk, S., & Vetterli, M. (2006). A frequency domain approach to registration of

aliased images with application to super-resolution. EURASIP Journal on Applied Signal Processing,

2006, 233.

[14] Sun, D., Roth, S., & Black, M. J. (2010). Secrets of optical flow estimation and their principles. In

Proceedings of IEEE conference on computer vision and pattern recognition (p. 2432–2439).

[15] Baker, S., & Kanade, T. (2002). Limits on super-resolution and how to break them. IEEE Transactions

on Pattern Analysis and Machine Intelligence, 24(9), 1167–1183.

[16] S. Baker and T. Kanade. Hallucinating faces. In Fourth International Conference on Automatic Face and Gesture Recognition, pages 83–88, 2000.

[17] Bee Lim, S. S. (2017). Enhanced Deep Residual Networks for Single Image Super-Resolution. Seoul, Korea: Department of ECE, ASRI, Seoul National University

[18] Chao Dong, C. C. (n.d.). Accelerating the Super-Resolution Convolutional Neural Network. Hong Kong: Department of Information Engineering, The Chinese University of Hong Kong.

[19] Enhanced Deep Residual Networks for Single Image Super-ResolutionBee LimSanghyun SonHeewon KimSeungjun NahKyoung Mu LeeDepartment of ECE, ASRI, Seoul National University, 08826, Seoul, Korea

[20] Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. TPAMI 38(2) (2015) 295–307

[21] Accelerating the Super-Resolution Convolutional Neural Network Chao Dong, Chen Change Loy, Xiaoou Tang

[22] J. Yang, J. Wright, T. Huang, and Y. Ma, “Image super-resolutionas sparse representation of raw image patches,” inIEEE Conferenceon Computer Vision and Pattern Recognition, 2008. 1, 3

[23] Fast and Accurate Image Super-Resolution withDeep Laplacian Pyramid NetworksWei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang

[24] Gunnar Farneback, (2003). Two-Frame Motion Estimation Based on Polynomial Expansion (p. 363-366)

[25] Christopher Zach, Thomas Pock, and Horst Bischof. A duality based approach for realtime tv-l 1 optical flow. In Pattern Recognition, pages 214–223. Springer, 2007.

[26] Thomas Brox, Andres Bruhn, Nils Papenberg, and Joachim Weickert. High accuracy optical flow estimation based on a theory for warping. In Computer Vision-ECCV 2004, pages 25–36. Springer, 2004.

[27] Jean-Yves Bouguet. Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm. Intel Corporation, 5, 2001