Super-Resolution in a Nutshell

M. A. de Jesus, Vania V. Estrela
Telecommunications Department, Universidade Federal Fluminense (UFF), RJ, Brazil
majesus1977br@gmail.com, vania.estrela.phd@ieee.org

Osamu Saotome
CTA-ITA-IEEA 12.228-900,
S. J. dos Campos, Brazil
osaotome@gmail.com

***Abstract*—SR approaches reconstruct a High-Resolution (HR) image from a set of Low-Resolution (LR) images, and restore an HR video from a group of neighboring LR frames. To overcome the limitation and/or ill-posed conditions of the image acquisition process as well as facilitating content visualization and scene recognition. This paper presents a comprehensive review of SR image and video reconstruction methods and their entailed challenges. Several SR applications are discussed along with practical applications in various real-world problems, like remote sensing, medical imaging, and biometric recognition. This survey groups works into different categories. For each group, basic concepts are studied. All-purpose concerns like imaging models, registration algorithms, optimization of the cost functions used, coping with color data, improvement factors, quality evaluation, and the most commonly employed databases are discussed.**

Keywords—super-resolution; image registration; fusion; image restoration; mosaicing; motion estimation

I. Introduction

Although monitoring cameras are ubiquitous, people's interest in details still calls for better pictures, due to limited equipment costs and constraints, weather conditions, as well as target shooting distances. Super-resolution (SR) refers to methods to upscale/upsize/restore pictures or video sequences. Starting from Low-Resolution (LR) images, SR recovers a High-Resolution (HR) one that provides better visual results with additive noise elimination, limited detector sizes, and optical constraints. It is related to image fusion, registration, and mosaicing. Some SR applications follows.

1) Frame freeze and Region of Interest (ROI) zoom for human perception and analysis.

2) Resolution enhancement for automatic target recognition.

3) In *surveillance*, several images from the same area can be used to render an improved resolution image.

4) In *medical imaging* (Ultrasound, CT, MRI, etc.) SR technique can be applied to enhance the resolution and build multi-modal versions of the area under study by combining data from several images acquired with limited resolution.

5) Video standard conversion, e.g. from PAL-M video signal to HDTV signal and from 4K to 8K.

SR approaches try to reproduce the process of losing quality when using LR cameras and then solve an ill-posed inverse problem of finding video, which being downsized with that process gives us known LR material, which does not have straightforward solution. This typically requires regularization, optimization, and extensive computational time. SR works effectively when several LR images contain slightly different perspectives of the same object. Then, all the object data exceed the knowledge present in a single frame. Motion Estimation (ME) can help to upscale an image. If an object is steady and appears identical in all frames, no additional knowledge is available. If there is movement or fast transformations, then the target will appear distincly in different frames, which requires more frames to reconstruct an HR frame.

This work is organized as follows: Section II, introduces the image degradation process, while Section III handles state-of-the-art SR. The challenges in SR image reconstruction are decribed in IV. Finally, conclusions are drawn in V.

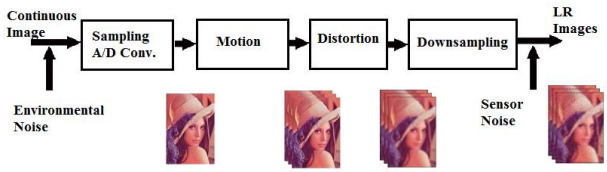

Fig. 1. Imaging degradation process.

II. The Image Degradation Model

The due to hardware limitations, the imaging system has imperfections, and various types of degradations as seen in Fig. 1. The Point Spread Function (PSF) models some kinds of optical and mechanical distortions. Since the image pixel is generated by integration over the sensor area, The limited sensor density leads to aliasing effects, limiting the image spatial resolution. These degradations are handled entirely or partly with different SR techniques.

This work uses the subsequent notation: (i) Upper case bold letters $\boldsymbol{X}$ and $\boldsymbol{Y}$ represent lexicographically ordered vectors for HR and LR frames. (ii) Lower case bold letters such as $\boldsymbol{x}$ and $\boldsymbol{y}$ stand for lexicographically ordered vectors for HR and LR image *patches* respectively. (iii) Underlined upper case bold letters show the result of a vector concatenation, e.g., is a vector concatenation of $\boldsymbol{Y}_k$, with $k=1,\ldots,K$, where K is the number of captured LR frames by the camera. Moreover, (iv) Plain upper case symbols denote matrices and simple lower case symbols refer to scalars.

Let $\boldsymbol{X}$ denote the desired HR image, i.e., the digital image sampled from the bandlimited continuous scene, and $\boldsymbol{Y}_k$ be the

k-th LR observation from the camera. There are K LR frame versions of $\boldsymbol{X}$, where each LR observation vector $\boldsymbol{Y_k}$ is related to the HR image $\boldsymbol{X}$ by

$$\boldsymbol{Y_k}=D_kH_kF_k\boldsymbol{X} + V_k, \qquad (1)$$

where F_k encodes the motion information for the *k-th* frame, H_k models the blurring effects, D_k is the downsampling operator, and V_k is the noise term. These linear equations can be rearranged into a large linear system

$$\underline{\boldsymbol{Y}}=\mathrm{M}\boldsymbol{X} + \underline{\mathrm{V}}. \qquad (2)$$

Matrices D_k, H_k, F_k, and M are very sparse, unknown, have to be estimated from the available LR observations, which worsens the linear system ill-conditioning. Thus, regularization is always advantageous and frequently essential. Next, some basic SR techniques proposed in the literature are examined.

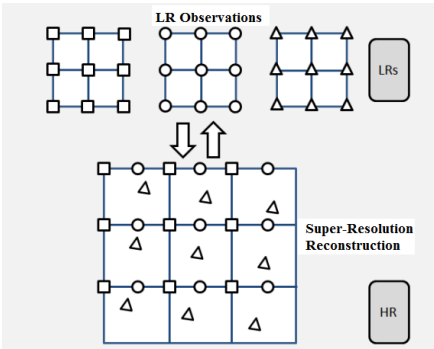

Fig. 2. The SR rationale.

III. State-of-the-Art Super-Resolution Techniques

SR reconstruction has been a very active research area and many techniques have been proposed to deal with this issue. Fig. 2. depicts the SR process.

A. Frequency Domain SR

Initial works on SR [A, B] explored Frequency Domain (FD) schemes with very restricted image observation models. Nowadays Spatial Domain (SD) methods are more common for their flexibility. They assume a global translation model with known parameters and noise free. Various extensions handling more complicated observation models have been proposed taking into account noise and blurring, with Tikohonov regularization and relying on ME. These approaches are computationally efficient, but limited due to more complicated image degradation models and to the use of various image priors within a regularization framework.

B. Interpolation-Restoration

SD approaches have been proposed to overcome the FD difficulties. The HR image and the LR frames are related as in (2). There are three stages for this approach 1) LR image registration; 2) non-uniform interpolation, and 3) deblurring and noise removal to get the desired $\boldsymbol{X}$. The LR frames are first aligned by some image registration algorithm to subpixel accuracy. These aligned LR frames are then put on a HR image grid, where non-uniform interpolation methods are used to fill in those missing pixels on the HR image. At last, deblurring is done by any classical deconvolutional algorithm with noise removal. These and other approaches are intuitive, simple and computationally efficient, assuming simple observation models. However, there is no optimality guarantee for the estimation. The registration error can easily propagate to the next stage and the interpolation step is suboptimal without noise and blur. Moreover, without an HR image prior as proper regularization, the interpolation based approaches need special treatment of limited observations to reduce aliasing.

C. Statistical Methods

Unlike the previous techniques, statistical approaches relate the SR reconstruction steps stochastically toward optimal reconstruction. The SR reconstruction problem is cast according to a full Bayesian framework by regarding the HR image and motions among LR inputs as stochastic variables, using a degradation matrix defined by the motion vector, and blurring kernel. Assuming a uniform prior over $\boldsymbol{X}$, reduces to the simplest Maximum Likelihood (ML) estimator (ME is assumed as a prior) and relies on the observations only, seeking the most likely solution for the observations to take place by maximizing $p(\underline{\boldsymbol{Y}}|\boldsymbol{X})$, giving

$$\hat{\boldsymbol{X}}_{ML} = (M^T M)^{(-1)} M^T \underline{\boldsymbol{Y}}. \qquad (3)$$

If $M^T M$ is singular, then regularization yields a unique solution, but computing $(M^T M)^{-1}$ is usually prohibitive due to high dimensions. A Back-Projection (BP) scheme can iteratively updates the current estimation by adding back the warped simulation error convolved with a BP Function (BPF). BPF was extended to real applications by incorporating multiple motions; deal with occlusion, transparency and some objects of interest. The BP algorithm is simple and flexible in handling many observations with different degradation procedures. However, its solution is not unique, depending on the initialization and the choice of BP kernel. The BP algorithm is an ML estimator and it implies some underlying assumptions about the noise covariance of the observed LR pixels. An ML SR image estimation algorithm robust to subpixel shifts, and image noise, that estimates the HR image simultaneously using the Expectation-Maximization (EM) algorithm and treats the MEs as unknown appeared in [c]. Since the ML estimator is usually very sensitive to noise, registration estimation errors and the PSF, proper regularization in the feasible solution space is always desirable, which leads to MAP strategies.

Many works in SR reconstruction followed Maximum a Posteriori (MAP) approaches, where the observation model assumptions and the prior term $Pr(\boldsymbol{X})$ for the desired solution vary. Commonly used image priors for the SR reconstruction techniques are: the Gaussian Markov Random Field (GMRF), the Huber MRF, and the Total Variation (TV) norm. Multiple frame SR reconstruction can be divided into two subproblems: LR registration and HR estimation. In Joint MAP (JMAP) estimation, ME and HR can benefit from each other with the motion and PSF estimates as unknowns. Registration, restoration and interpolation are solved simultaneously by maximizing the likelihood using EM and be applied to compressed video. For more complex multiple moving objects in an SR setting, the problem can be solved by combining MAP with ME, segmentation and SR reconstruction. Overfitting may be an issue.

Simple models spanned by few parameters may be sufficient for SR some applications. Given the LR observations, estimating parameters by integrating over the unknown HR image (Bayesian treatment) is a useful approach [96] where the marginal is used to estimate the PSF and motion parameters. To ease the problem analytically, a Gaussian Process Prior (GMRF) is used to model the HR image. Such a Bayesian approach outperforms JMAP approaches, which will easily get overfitting with the PSF parameters. The integration over the HR image is computationally intensive, and the Gaussian prior over the image may add excessive smoothness. Instead of marginalization over the unknown HR image, the unknown PSF and motion parameters can be integrated to overcome the uncertainty of the registration parameters [79]. The registration parameters are estimated beforehand and then treated as Gaussian variables with the pre-estimated values to model their uncertainty. Although Bayesian treatments marginalizing the unknowns are promising, image priors or registration parameters have to take simple parametric forms, which limits applicability in real videos. Computation could also be a concern.

D. Computational and Cognitive Strategies

Generic image priors help regularize the solution properly, but are not sufficient. The regularization becomes especially crucial with an insufficient number of measurements, as when only one LR frame is observed [2]. Recently, Example-Based (EB) methods helped regularizing the SR reconstruction problem, and to break the SR limit caused by inadequate measurements. EB methods develop the prior by sampling other images in a local way, and are effective when insufficient observations are available. Some issues must be addressed: (i) the choice the optimal patch size given the target image; (ii) different databases with diverse statistics are necessary; (iii) how to use the EB prior more efficiently.

Projection onto Convex Sets (POCS) formulates the SR problem as multiple constraining convex sets containing the desired image within the sets. A convex set is flexible and can incorporate different kinds of constraints or priors, even nonlinear and nonparametric constraints. It can handle motion blur [69], but it may require post-processing. The POCS can be further extended for robust, object-based SR reconstruction. The advantage of the POCS technique easily incorporates any kinds of constraints and priors difficult for stochastic approaches. POCS drawbacks are the heavy computation, slow convergence, and non-unique solution depending on the initial guess, the need for priors on the motion parameters and system blurs, and incapacity to estimate parameters simultaneously.

To some extent, the proper model formulation is the basis for finding the solution. Computational intelligence offers an alternative to optimization that has a simple implementation, higher computational accuracy, and lower time cost in demanding applications. Experimental results using evaluation metrics show that these methods can improve both objective and subjective results. Other possibilities of solving the SR problem may use machine learning and compressive sensing.

IV. Bottlenecks

SR uses subpixel accurate MC to find similar areas in neighbor frames to intelligently merge them combining the details. There are cases when this cannot be done successfully or cannot bring new details.

Absence of Change: Without any movement or change, the target is identical among several frames, there is little information to extract, and the quality will be as of simple spatial upsize.

Abrupt Motion: Sudden motion is difficult to track leading to low-quality results due to the following reasons: (i) motion blur caused by the camera exposure; and (ii) compression using delta-frames, because intense motion creates many differences between frames (the codec quantizes data strongly to fit into required bitrate, so more details are lost on the way).

Heavy Compression: If your video is compressed to a low bitrate, in many cases this is very bad for SR. There are two types of lossy video codecs: those that use delta-frames and those that use only key-frames. If the video is compressed heavily by a key-frame-only codec like MJPEG (DV is also such codec, but DV video always has high enough bitrate), then each frame is compressed independently and a lot of details are lost in each frame. This generally leads to blocking artifacts, which are common in heavily compressed JPEG images because the object changes a lot during motion. Hence, to precisely track motion and get some details became impossible. Using other frames results in blocking artifacts that worse picture quality. Since as MC, SR uses differences between frames, then it will find only gross changes ineffective to improve the frame (loss of details).

Image Registration: It is critical for the success of SR reconstruction, where complementary spatial samplings of the HR image are fused, besides being an ill-posed procedure. Having LR observations heavy aliasing artifacts worsens the process. The performance of standard image registration methods decreases as the resolution of the observations is

reduced, resulting in more registration errors. LR picture registration and HR image estimation are essentially interdependent processes. Accurate subpixel ME improves HR image estimation. Conversely, high-quality HR image can facilitate accurate ME. So, the LR image registration can be tackled with the HR image reconstruction, leading to a joint ML or MAP framework for simultaneous estimation, which captures the dependence between LR image registration and HR image estimation. Still, with limited observations, the joint estimation of registration parameters and HR image may result in overfitting. Stochastic approaches dealing with the HR image estimation and image registration simultaneously are promising; however, the motion models can limit performance. Optical flow ME can be applied to scenarios that are more intricate. However, the insufficient measurements for local motion estimations make SR algorithms vulnerable to errors. The 3-D SR problem brings in several extra challenges.

Computation Efficiency: Intensive calculations due to the large number of unknowns required by expensive matrix manipulations hinder real-time implementations. The number of calculations goes up significant for non-translation models, which can be ameliorated by massive parallel computing. The FPGA technology can facilitate real-time SR systems. For videos with arbitrary motions, promising results come from parallel computing, such as GPUs, and other special hardware implementations.

Robustness: Traditional SR techniques are vulnerable to the presence of outliers due to motion errors, noise, inaccurate blur models, moving objects, and so forth. These inaccuracies cannot be treated as Gaussian noise, which is the usual assumption when the l2-normis used. SR robustness is important because the parameters describing the image degradation model cannot be estimated perfectly, and sensitivity to outliers may cause visually distressing artifacts.

Performance Limits: They will shed light on SR camera design, helping to investigate factors like model errors, the number of frames, and zooming factors, but an ambitious analysis of the performance limits for all SR techniques can be complicated. First, SR reconstruction is a difficult task consisting of many interdependent components. Second, it is still unknown what is the most informative prior given the SR task. Last, good metrics are still needed. It is desirable to extend models with a detailed analysis of SR performance with factors such as ME, decimation factor, the amount of frames, and prior information. While it is hard to draw consistent conclusions for different SR techniques regarding performance evaluation, some benchmarks, and realistic datasets can ease algorithm comparison and understanding.

V. Conclusion

This paper reviews SR image and video reconstruction methods and points out research challenges. SR image approaches reconstruct a single HR image from a set of LR images, and the SR video techniques recreate an HR image sequence with a group of adjacent LR frames. Moreover, several SR applications are considered along with some insightful remarks on future SR research directions. Note that SR cannot always provide excellent results but SR reconstruction techniques may include pre-processing and post-processing. The image post-processing stage helps to validate the resolution.

This survey provides an SR overview with a loose, broad taxonomy. For each taxonomic group, the core algorithmic concepts are examined, and common issues such as imaging models, registration schemes, cost functions optimization employed, color handling, improvement factors, metrics for quality assessment, and adequate databases for testing and validation are discussed.

References

[1] R. Y. Tsai and T. S. Huang, "Multiframe image restoration and registration," in Advances in Computer Vision and Image Processing, vol. 1, chap. 7, pp. 317-339, JAI Press, Greenwich, Conn, USA, 1984.

[2] M. Irani and S. Peleg, "Super resolution from image sequences," in Proc. of the 1991 ICPR, vol. 2, 1990, pp. 115-120.

[3] http://www.infognition.com/articles/ Retrived on November 21, 2016.

[4] R. Y. Tsai and T. S. Huang. Multipleframe image restoration and registration. In *Advances in Computer Vision and Image Processing*, pages 317-339. Greenwich, CT: JAI Press Inc., 1984.

[5] B. C. Tom, A. K. Katsaggelos, and N. P. Galatsanos. Reconstruction of a high resolution image from registration and restoration of low resolution images. In *Proc. of IEEE International Conf. on Image Processing*, pages 553-557, 1994.

[6] [3] A. M. Coelho, V. V. Estrela, F. P. do Carmo, & S.R. Fernandes, Error Concealment by Means of Motion Refinement and Regularized Bregman Divergence, Proc. of the 13th Int'l Conf. on Int. Data Eng. and Automated Learning-IDEAL'12, pp. 650-657, Natal, Brazil, 2012, Springer-Verlag. doi: 10.1007/978-3-642-32639-4_78

[7] [4] S. R. Fernandes, V. V. Estrela, H. A. Magalhaes, & O. Saotome, On Improving Sub-Pixel Accuracy by Means of B-Spline, Proc. of the 2014 IEEE Int'l Conf. on Imaging Syst. and Techniques (IST 2014), Santorini, Greece. doi: 10.1109/IST.2014.6958448

[8] Singh, P., and Sharma, J. (2016). A review paper on image denoising by low rank matrix decomposition and genetic algorithm. International Journal of Engineering Sciences & Research Technology, 5(7), 166–681. http://doi.org/10.5281/zenodo.56923

[9] Irani, M., Peleg, S.: Motion analysis for image enhancement: resolution, occlusion, and transparency. J. Vis. Commun. Image Represent. 4, 324–335 (1993)

[10] Altunbasak, Y., Patti, A.J., Mersereau, R.M.: Super-resolution still and video reconstruction from mpeg-coded video. IEEE Trans. Circuits Syst. Video Technol. 12, 217–226 (2002)

[11] Tom, B.C., Galatsanos, N.P., Katsaggelos, A.K.: Reconstruction of a high resolution image from multiple low resolution images. In: Chaudhuri, S. (ed.) Super-Resolution Imaging. Kluwer, Norwell (2002).

[12] Reibman, A.R., Bell, R.M., Gray, S.: Quality assessment for super-resolution image enhancement. In: Proc. of IEEE International Conf. on Image Processing, USA, pp. 2017–2020 (2006).

[13] Robinson, D., Milanfar, P.: Statistical performance analysis of super-resolution. IEEE Trans. Image Process. 15(6), 1413–1428 (2006)

[14] Brandi, F., de Queiroz, R.L., Mukherjee, D.: Super-resolution of video using key frames and motion estimation. In: Proc. of IEEE Int'l Conf. on Image Processing, USA, pp. 321–324 (2008).

[15] Costa, G.H., Bermudez, J.: Registration errors: are they always bad for super-resolution? IEEE Trans. Signal Process. 57(10), 3815–3826 (2009)

[16] Yang, J., Wright, J., Huang, T., Ma, Y.: Image super-resolution via sparse representation. IEEE Trans. Image Process. 19(11), 2861–2873 (2010)