SURECAVI : Super-resolution for Camera system in Visible domain

Andrea Petreto[1-3], **Gabriele Facciolo**[2], **Thomas Romera**[1-3], **Boris Gaillard**[1], **Thomas Eboli**[2], **Pablo Arias**[2], **Jérémy Anger**[2], **Thibaud Ehret**[2], **Ngoc Long Nguyen**[2], **Miguel Colom**[2], **Pascal Dupuy**[1], **Patrice Ménard**[1], and **Jean-Michel Morel**[2]

[1]*LERITY, 10 av de l'entreprise, 95862, Cergy-Pontoise, France, Email: apetreto@lerity-alcen.com*
[2]*Université Paris-Saclay, CNRS, ENS Paris-Saclay, Centre Borelli, 91190, Gif-sur-Yvette, France*
[3]*Université, CNRS, LIP6, F-75005 Paris, France*

KEYWORDS:

Super-resolution, Video-enhancement, real-time processing.

ABSTRACT:

Video Enhancement and image quality improvement are major topics in surveillance and defense domains. The goal is to always push limits of Detection, Recognition and Identification (DRI). One solution consists in applying software algorithms for video enhancement. Here we focus on visible domain applications and on the identification part.

For degraded visibility situations, video quality can rapidly decrease. Therefore it can become difficult for an operator to take a fully informed decision. In this paper we tackle the problem of real-time super-resolution and video deblurring. Although many algorithms have shown impressive results in terms of image quality, they usually suffer from very high computational complexity.

In this paper we introduce a new project of a visible camera system that will be able to perform live super resolution-and deblurring. This project relies on the previous work performed by LERITY and Centre Borelli (CB) including 3 PhD thesis funded by DGA (French ministry of defense).

1. INTRODUCTION

Better video quality is a major topic in defense and security area. Here we focus on visible domain applications, which are well known for providing the best identification range. For those systems, threat identification rely on high quality images so the operator can take a fully informed decision as soon as possible.

However, long range observation, low light conditions or other degraded visibility situations can rabidly decrease the video quality: noise, blur, turbulence's distortion, lack of detail/contrast...

In this paper we introduce the SURECAVI project: "SUper REsolution for CAmera system in VIsible domain". This project stands as part of a collaboration between LERITY ans Centre Borelli (CB) and is following 3 PhD thesis on super-resolution and deblurring and on image processing optimization for embedded systems.

LERITY is a player in the field of optronics, specialized in very high definition vision systems in the visible spectrum, operating in particular in degraded visibility conditions [1]. Among the products designed by LERITY, the improvement brought by the new algorithms concerns systems integrating remote resources such as CATEYE System and CATEYE System XLR.

The team of Centre Borelli (CB) ENS Paris Saclay has developed a fast super-resolution pipeline to process series of push-frame satellite images from the PLANET SKYSAT satellites [2]. This pipeline is currently being industrialized by KAYRROS, and enables the efficient production of images with a resolution higher than those provided by PLANET.

We detail the hardware and algorithmic adaptations necessary to bring these advances to CATEYE System. The main hardware adaptations concern the optimization of the Point Spread Function (PSF) of the optics, the generation and the control of the sensor micro displacements.

The goal of the project is to significantly improve the resolution and image quality of these cameras through software and hardware improvements, at (almost) the same hardware price.

The project will rely in particular on super-resolution techniques by merging several raw images acquired by the camera. The image produced will be super-resolved and its image quality will be improved by real-time deblurring and denoising. All of these operations require an increasing computing power. But thanks to new generations of GPUs and algorithms that take advantage of the massive parallelism, it is now possible to perform these operations in real time. In addition, CB and LERITY dispose of a wide range of algorithms of increasing complexity enabling the adaptation of a processing chain to the computational constraints.

The expected result at the end of the project is a prototype of a super-resolved visible observation system and a real-time processing chain with a TRL of 6.

In this paper we first present the previous work performed by LERITY and CB that will serve as a foundation for the SURECAVI project. Then, we will detail our future work and the main challenges we plan to encounter.

2. PREVIOUS WORK

2.1. CB's previous work on deblurring and super resolution

CB's team has acquired a solid experience in image and video processing. They notably invented major denoising algorithms like NL-means [3], NL-Bayes [4], and VNLBayes [5, 6]. Through the PhD thesis of J. Anger, they have also recently acquired better skills in video deblurring and super resolution with significant contributions to the state-of-the-art [7, 7–9]. This thesis tackles the blur issue and its suppression according three situations: first to restore images burst with super-resolution and deblurring, then to non-blind and blind deblurring. The algorithms introduced during this PhD thesis [10] have been applied in a processing chain for SKYSAT images of PLANET, which delivers higher definition than the previously used method. Fig. 1 shows the performance improvement on a real target compared to the previous SKYSAT method.

CB's team, through J. Anger's thesis, studied and compared multi-frame deblurring and super-resolution algorithms. They started with Fourier Burst Accumulation [11], which efficiently merges images across time through a weighted average in the Fourier domain. CB adapted this method for video in [12]. The thesis then focuses on multi-frame super-resolution methods. It shows that recent progress in *push-frame* satellite conception allows for increased spatial resolution via multi-frame super-resolution algorithms.

Therefore, a fast multi-frame super-resolution method based on a high order spline interpolation has been proposed. It combines several low resolution images to produce a high resolution output. Like most multi-frame based algorithms, the proposed method uses global movement compensation and merging. The compensation step estimates affine transformations to a sub-pixel level between low definition images. Using this precise information, the merging step combines all samples within one high definition image using an adaptation of the ACT algorithm [13]. Instead of trigonometric polynomial interpolation, this algorithm uses spline interpolation.

The formation model of the low resolution image v_i can be described as in Eq. 1.

$$v_i = \Sigma_1((u \circ A_i) * k) + n_i, \tag{1}$$

where A_i is a projective transformation and k is a convolution kernel involving all the effects similar to the PSF (eg. optical blur and pixel integration). The σ_1 operator is the bi-dimensional sampling operator and n_i represents the image noise.

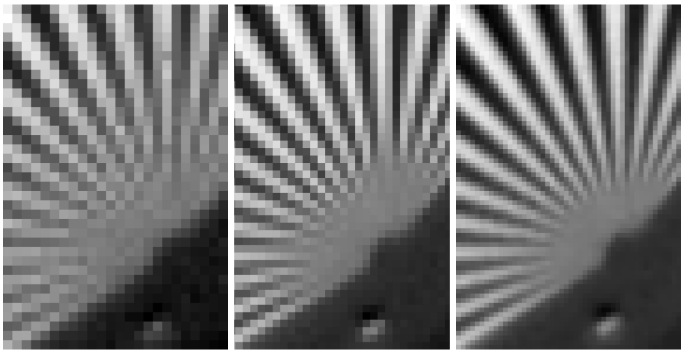
Synthetic and real-world experiments (cf. Fig. 1) showed that the proposed method brings a resolution gain of 10 cm per pixel compared to the previous method (53cm/pixel instead of 63cm/pixel starting from 90cm/pixel images). This is a starting point for the SURECAVI project.

Another topic of the thesis was to study non-blind deblurring. It is shown that most methods assume a conventional but too simple image formation model: $(v := u * k + n)$, where only noise and blur kernel are considered. This lead to study in detail a more realistic image acquisition chain that explicitly models saturation, pixel quantification and the non-linear response of the camera. The thesis proposed a variational method which performs better than the state-of-the-art and provides good visual results by eliminating ringing artifacts around saturated zones. This method is published in [14].

The third topic addressed by J. Anger's PhD thesis was to make blind deblurring robust to noise. In the considered set up, the blurring kernel is unknown and must be estimated. The goal is to retrieve sharp images without knowing the camera movement nor the blur induced by the sensor. Two representative methods of the state-of-the-art where studied. First the Goldstein and Fattal method [15]. It estimates the blur directly from statistic irregularities within the power spectrum of blur images. The other state-of-the-art solution is a variational method that uses an *a priori* ℓ^0 [16] on the images gradients. It estimates the kernel by iterating between a sharp image estimation and the kernel.

Even though recent methods provide very good results [16–19] on non-noisy images (or very low noise levels), they fail to estimate the kernel on noisy situations. However, the hypothesis of non-noisy blurred image in real life is not realistic since blur and noise usually appear under the same conditions.

To tackle this issue the thesis proposed to improve the ℓ^0 method [16]. The goal is to handle better high levels of noise [7]. The key idea is to observe that the kernel estimation is highly affected by the noise in the input images. The thesis demonstrated that it is very efficient to denoise the image before the deconvolution. Although very simple as a concept this solution requires an iterative denoising method which is very demanding in terms of computation time. A comparison between the proposed and other state-of-the-art methods is shown in Fig. 2.



(a) Low resolution input **(b)** SKYSAT L1B **(c)** Proposed solution

Figure 1: Restoration example of a real target from 35 low resolution frames. From left to right: reference low resolution frame, L1B output provided by Planet (zoom factor $\times 1,25$), proposed method output (zoom factor $\times 2$).

In the context of super resolution, it is mandatory to apply a sharpening step to adapt the system PSF. The studied and developed algorithms allow to efficiently process images while taking into account the effect of noise. A GPU implementation should significantly improve the computation time. In the case of small blur kernels, approximated but very fast methods [20] will also be considered.

2.2. LERITY's background and previous work

LERITY's specialty is to develop high performance cameras for long range vision and degraded visibility conditions. Among the high performance products designed by LERITY stands the CATEYE family. The different versions of the CATEYE family are presented in Fig. 3.

The first *CATEYE* is a portable long range camera with active imaging. *CATEYE Space* is its specialized version. It is designed for space debris surveillance.

CATEYE System is a long range surveillance system mounted on a motorized turret. It is composed of one wide angle color camera and one long range highly sensitive black and withe camera. The visualization, the camera and turret control are then performed by the operator via a distant computer. This very new product is already used for French coast surveillance in Normandy and could be deployed on many other locations. LERITY is currently working on its illuminated version *CATEYE System XLR*. All of these systems perform in visible domain in order to provide the best identification range possible.

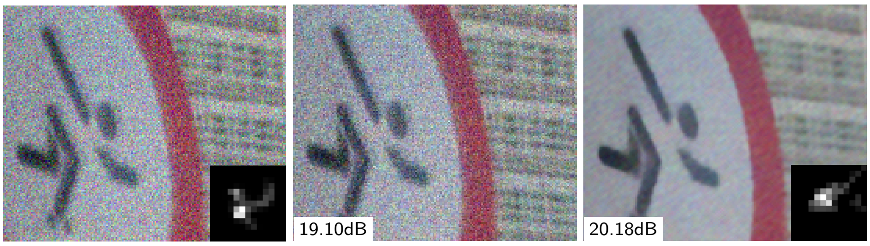
Regarding the computational complexity of the deblurring and super-resolution algorithms we wish to implement, only *CATEYE System* and *CATEYE System XLR* will be considered for the SURECAVI project. Those two systems indeed possess a deported processing unit with less power consumption constraints than the regular *CATEYE*.

Concurrently LERITY has also acquired a solid experience in algorithmic and code optimization for embedded image processing. It notably founded 3 PhD theses in collaboration with the Computer Science laboratory LIP6. The first one, led by A. Petreto, was on real-time embedded video denoising. The other two are still in progress; one is led by T. Romera and tackles optical flow optimization for embedded GPU. The other one led by M. Millet covers the problem of CPU optimization of irregular algorithms for embedded systems.

Through these works LERITY notably provided a new real-time video denoising algorithm called RTE-VD [21]. Compared to other state-of-the-art methods, RTE-VD brings a new trade-off between speed and denoising quality [22]. It was successfully implemented on the real-time embedded video denoiser VIRTANS [23], for live Full HD denoising (1920×1080 pixels). VIRTANS is a 10cm×7cm×6cm video denoiser visible in Fig. 4.

Figure 4: VIRTANS prototype: VIdeo Real-Time Algorithm: Noise Supression.

VIRTANS is able to denoise SDI Full HD video with RTE-VD and display the result through HDMI port. The power consumption is maintained under 10 Watts. Compared to other state-of-the-art methods, RTE-VD brings a new trade-off between denoising quality and computation speed. A visual comparison is shown in Fig. 5. Unlike the other real-time methods RTE-VD remains efficient under very noisy situations while it's computation time is one order of magnitude lower than reference methods for quality.



(a) Input ($\sigma = 10\%$) **(b)** Tao [17] **(c)** Zhong [18]

(d) Zhou [19] **(e)** Pan [16] **(f)** Proposed

Figure 2: Blind deblurring under high noise level and comparison with other state-of-the-art methods. The proposed method is able to estimate the kernel and restore high quality image despite the presence of noise.

Figure 3: CATEYE family.

Along RTE-VD, LERITY notably provided a highly optimized implementation of TV-L1 optical flow estimation on both CPU [21, 24] and GPU [25]. To our knowledge these are the fastest implementations on both architectures. This was made possible by using High level transforms and optimizations techniques [26] like operator fusion, operator pipeline, SIMD and MIMD parallelization...

Fig. 6 shows the processing speed of different implementations of TV-L1 on the Nvidia's Jetson AGX embedded platform, namely:

- OpenCV XX: F_{32} OpenCV implementation on the XX architecture (CPU or GPU)
- XX_neon CPU: LERITY'S fully optimized CPU version using XX (F_{32} or F_{16}) computation format,
- XX_base: GPU baseline version using XX format
- XX_global: LERITY's Intermediate GPU implementation using XX format
- XX_shared_fusion: LERITY's GPU fully optimized implementation with operator fusion using shared memory and XX format.

The timing results in Fig. 6 shows that the optimized CPU versions are faster than the f32_base GPU version. The f32_neon CPU version is also faster than the OpenCV GPU version for images less than 400×400 pixels and the f16_neon CPU version is faster for images less than 800×800. Our f16x2_shared_fusion GPU version is $3\times$ faster than OpenCV GPU and $7\times$ faster than f32_base. Better data reuse, fewer reads and writes to global memory, along with smaller data size due to the use of F_{16} lead to a lower memory footprint of our implementations.

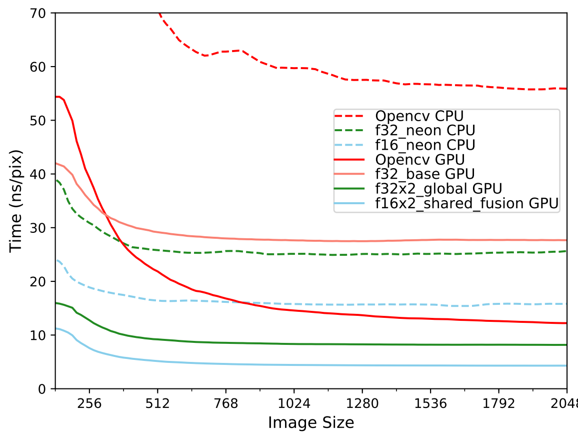


Figure 6: Execution time of TV-L1 optical flow estimation in nanoseconds per pixel depending on the image size on both CPU and GPU of the Nvidia's AGX platform. Comparison between LERITY's implementations and non optimized (base) and OpenCV implementations.

Figure 5: Denoising comparison between RTE-VD and other state-of-the-art methods. *Noisy* and *Origin* are the same since the input is naturally noisy. On the same CPU architecture, RTE-VD is $2.2\times$ slower than STMKF and respectively $200\times$ and $4600\times$ faster than VBM3D and VBM4D.

The goal of the SURECAVI project is to take full advantage of the algorithmic skills of CB and the expertise of LERITY in optronics and embedded processing to provide a new super resolved vision system. The enhanced image should be played to the operator in real-time.

3. FUTURE WORK: SURECAVI PROJECT

The SURECAVI project will be based on the previous work performed by both CB and LERITY that we presented in the previous section. Starting from a *CATEYE System* platform, we will improve it by implementing a highly efficient, real-time super-resolution an deblurring solution. This will probably require resizing of the deported processing unit as well as changes of the optical set to be suitable for super-resolution. CB's algorithms will have to be adapted to the real-time constraints.

The project will be split in three main steps. The first one will be carried by LERITY and concerns the prototype development for super-resolution. It includes preliminary studies, hardware modifications and HMI integration. The second step will be carried by CB along with the first one. It concerns the processing chain adaptations to fit with the new system constraints. The third step will be a common effort to integrate, fine tune and validate the algorithms into *CATEYE System* and *CATEYE System XLR*.

3.1. Prototype development for super-resolution

This step will be carried by LERITY and mainly consists into acquiring representative video and hardware specifications and modifications of the prototype. LERITY will acquire baseline sequences with an existing *CATEYE System* device. It will also characterize the optical geometry of the system in order to evaluate the needed modifications suitable for super-resolution. Moreover sequences will also be captured from cameras with excellent PSF to compare with the existing system. A good PSF is mandatory to fully exploit super-resolution techniques. LERITY will provide all these information to CB so they can adapt their algorithms.

In addition LERITY will develop and perform hardware modifications on the prototype to fit with super-resolution and deblurring constraints. It includes: the modifications of the optics, the integration of a color sensor (for long range), the integration of a piezo for micro displacements (needed for super resolution), the adaptation of the input data flow to allow extraction of data with a high dynamic range, the HMI modifications and the sizing of the processing unit.

3.2. Modeling and realization of algorithms without access to the prototype

This step will be carried by CB. CB will focus on adapting their algorithms to the standard sequences provided by LERITY. First they will rapidly provide an offline processing chain adapted to the specificities of LERITY's camera and their applications. They will be able to characterize the specific defaults within the images: Noise, distortion, blur...). This will give a way to evaluate CB's algorithms without the proper prototype. The developed algorithms will then be incorporated in a complete processing chain with fast image registration, mobile objects detection, denoising, super resolution and deblurring. A second processing chain that handles color mosaic will also be proposed. Color mosaicing presents specific challenges like color aliasing or demosaicing policy (before or after denoising).

These processing chains will be evaluated on the sequences provided by LERITY and also on simulated ones. Finally, CB will be able to provide multiple solutions with various computational complexities. The chosen solutions will be transmitted to LERITY for fine tuning and integration in the SURECAVI prototype.

3.3. Final integration and validation

This last step will be jointly carried out by both CB and LERITY Once the two first steps realized the chosen algorithmic solutions will be integrated within the new prototype. A first step will probably require implementation optimizations from LERITY to reduce computation time as much as possible. Then the solution will be evaluated: the gain in terms of performance will be particularly examined but also the ergonomics. Once the final integration complete, a global validation will evaluate the DRI improvements compared to the previous *CATEYE System*.

ACKNOWLEDGEMENT

This work will be partially funded by the *Direction générale de l'armement* (DGA), french Ministry of Armed Forces.

References

[1] LERITY-ALCEN, "Lerity's website," 2021, url = https://www.lerity-alcen.com/fr.

[2] Planet, "Skysat," 2021, url = https://www.planet.com/products/hi-res-monitoring/.

[3] A. Buades, B. Coll, and J. Morel, "Image enhancement by non-local reverse heat equation," *Preprint CMLA*, 2006. [Online]. Available: http://dmi.uib.es/{~}tami/publicacions/CMLA2006-22.pdf

[4] M. Lebrun, A. Buades, and J.-M. Morel, "A non-local bayesian image denoising algorithm," *SIIMS*, 2013.

[5] P. Arias and J.-M. Morel, "Video denoising via empirical bayesian estimation of space-time patches," *JMIV*, vol. 60, no. 1, pp. 70–93, Jan 2018. [Online]. Available: https://doi.org/10.1007/s10851-017-0742-4

[6] P. Arias, G. Facciolo, and J.-M. Morel, "A comparison of patch-based models in video denoising," in *IEEE Image, Video, and Multidimensional Signal Processing Workshop*. IEEE, 2018.

[7] J. Anger, M. Delbracio, and G. Facciolo, "Efficient blind deblurring under high noise levels," in *IEEE ISPA*, 2019, pp. 123–128.

[8] J. Anger, C. de Franchis, and G. Facciolo, "Assessing the sharpness of satellite images: Study of the planetscope constellation," in *IEEE IGARSS*. IEEE, 2019, pp. 389–392.

[9] J. Anger, G. Facciolo, and M. Delbracio, "Blind image deblurring using the l0 gradient prior," *Image Processing On Line*, vol. 9, pp. 124–142, 2019.

[10] J. Anger, T. Ehret, C. de Franchis, and G. Facciolo, "Fast and accurate multi-frame super-resolution of satellite images," *ISPRS Annals*, vol. 1, pp. 57–64, 2020.

[11] M. Delbracio and G. Sapiro, "Removing camera shake via weighted fourier burst accumulation," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3293–3307, 2015.

[12] J. Anger, "Défloutage de vidéos en temps réel," in *ORASIS 2017*, Jun. 2017. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01866678

[13] T. Strohmer, "On discrete band-limited signal extrapolation," *Contemporary Mathematics*, 1995.

[14] J. Anger, G. Facciolo, and M. Delbracio, "Modeling realistic degradations in non-blind deconvolution," in *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018, pp. 978–982.

[15] ——, "Estimating an Image's Blur Kernel Using Natural Image Statistics, and Deblurring it: An Analysis of the Goldstein-Fattal Method," *IPOL*, vol. 8, pp. 282–304, 2018.

[16] J. Pan, Z. Hu, Z. Su, and M.-h. Yang, "Deblurring Text Images via L0-Regularized Intensity and Gradient Prior," *Cvpr*, pp. 2901–2908, 2014. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2014/papers/Pan_Deblurring_Text_Images_2014_CVPR_paper.pdf

[17] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia, "Scale-recurrent network for deep image deblurring," 2018.

[18] L. Zhong, S. Cho, D. Metaxas, S. Paris, and J. Wang, "Handling noise in single image deblurring using directional filters," *IEEE CVPR*, pp. 612–619, 2013.

[19] X. Zhou, M. Vega, F. Zhou, R. Molina, and A. K. Katsaggelos, "Fast Bayesian blind deconvolution with Huber Super Gaussian priors," *Digital Signal Processing: A Review Journal*, vol. 60, pp. 122–133, 2017.

[20] M. Delbracio, I. Garcia-Dorado, S. Choi, D. Kelly, and P. Milanfar, "Polyblur: Removing mild blur by polynomial reblurring," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 837–848, 2021.

[21] A. Petreto, T. Romera, F. Lemaitre, I. Masliah, B. Gaillard, M. Bouyer, Q. L. Meunier, and L. Lacassagne, “A new real-time embedded video denoising algorithm,” in *IEEE International Conference on Design and Architectures for Signal and Image Processing (DASIP)*, 2019, pp. 1–6.

[22] A. Petreto, “Débruitage vidéo temps réel pour systèmes embarqués,” Ph.D. dissertation, Sorbonne université, 2020.

[23] A. Petreto, T. Romera, F. Lemaitre, M. Bouyer, B. Gaillard, P. Menard, Q. Meunier, and L. Lacassagne, “Real-time embedded video denoiser prototype,” in *9th International Symposium-Optronics in Defense and Security (Optro)*, 2020.

[24] A. Petreto, A. Hennequin, T. Koehler, T. Romera, Y. Fargeix, B. Gaillard, M. Bouyer, Q. L. Meunier, and L. Lacassagne, “Energy and execution time comparison of optical flow algorithms on SIMD and GPU architectures,” in *2018 Conference on Design and Architectures for Signal and Image Processing (DASIP)*. IEEE, 2018, pp. 25–30.

[25] T. Romera, A. Petreto, F. Lemaitre, M. Bouyer, Q. Meunier, and L. Lacassagne, “Implementations impact on iterative image processing for embedded gpu,” in *European Signal Processing Conference (EUSIPCO)*, 2021.

[26] L. Lacassagne, D. Etiemble, A. Hassan Zahraee, A. Dominguez, and P. Vezolle, “High level transforms for SIMD and low-level computer vision algorithms,” in *Proceedings of the 2014 Workshop on Programming models for SIMD/Vector processing (WPMVP)*, 2014, pp. 49–56.