# PhD Thesis Proposal

**Title:** Bridging the Gap Between Perceptual Quality and Visual Recognition in the Smart Video Surveillance System

## Brief description of the topic

During the past decade video surveillance systems have evolved from simple video acquisition and display systems to intelligent autonomous systems, capable of performing complex procedures. Nowadays, a video surveillance system can integrate some of the most sophisticated image and video analysis algorithms from research areas such as classification, pattern recognition, decision-making, image enhancement and other related topics. Thus, a modern surveillance system consists of different modules, namely image and video acquisition devices, data processing and analysis and coding and transmission systems. All these components are crucial for scene analysis and understanding for decision making and monitoring taks. Recently, with the renewed interest in artificial intelligence and the development of intelligent sensors technologies as well as high performance computing resources, we are witnessing a proliferation of deep learning-based solutions. Most of the developed algorithms for scene analysis and interpretation are based on deep neural networks (DNNs).

Indeed, even if we are far from being able to model and justify in a convincing and explainable mathematical framework the use of DNN-based approaches, it remains that this trend is really supported by the efficient solutions developed so far for the resolution of various problems and in particular in the field of computer vision and robotics, e.g., face recognition, pedestrian detection and pose estimation [1,2]. For many computer vision problems, the DNNs are trained and validated based on the assumption that the input images/videos are pristine (i.e., artifact-free) [3,4]. However, in video surveillance systems different signal processing stages like video capture, network transmission and video compression can affect the acquired video quality, thereby causing video distortions. Moreover, such systems are often deployed in outdoor structures such as stadium, buildings, parking and streets etc. Based on the nature of outdoor environment, some natural degradation due to adverse weather conditions such as haze, fog and smoke in some cases greatly reduce the visual quality of outdoor surveillance videos. These factors reduce the event/object detection and recognition efficiency to a great extent and hence are a cause of concern in the context of safety and security [5].

While there have been important advances in the area of computational photography to restore or enhance the visual quality of an image, the capabilities of such techniques have not always translated in a useful way to visual recognition tasks [6]. Consequently, there is a pressing need for the development of algorithms that are designed for the joint problem of improving visual appearance and recognition [7], which will be an enabling factor for the deployment of visual recognition tools in many video surveillance systems.

In this thesis, we will introduce a new dataset as a large-scale benchmark composed of video imagery captured under challenging conditions, and different enhancement tasks designed to test algorithmic impact on visual quality and automatic object recognition. Furthermore, we will propose a set of metrics to evaluate the joint improvement of such tasks as well as individual algorithmic advances,

including a novel objective metric for human observer and a set of quantitative measures for object recognition performance.

Using this set of metrics, we will present an in-depth analysis of these algorithms and a host of deep learning-based and classic baseline approaches. By building a bridge between computational photography and visual recognition, we will increase the robustness of DNNS that will provide promising prediction performance for images/videos with a broad range of quality levels.

**References**

[1] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,” in IEEE ICCV, 2015.
[2] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, “Deepface: Closing the gap to human-level performance in face verification,” in IEEE CVPR, 2014.
[3] Dodge, Samuel, and Lina Karam. "A study and comparison of human and deep learning recognition performance under visual distortions." In 2017 26th international conference on computer communication and networks (ICCCN), pp. 1-7. IEEE, 2017.
[4] Chen, Zhuo, Weisi Lin, Shiqi Wang, Long Xu, and Leida Li. "Image quality assessment guided deep neural networks training." arXiv preprint arXiv:1708.03880 2017.
[5] A. Beghdadi, M. Asim, N. Almaadeed, and M. A. Qureshi, “Towards the design of smart video-surveillance system,” in 2018 NASA/ESA Conference on Adaptive Hardware and Systems (AHS). IEEE, 2018, pp. 162–167.
[6] B. RichardWebster, S. E. Anthony, and W. J. Scheirer, “Psyphy: A psychophysics driven evaluation framework for visual recognition,” IEEE T-PAMI, 2018,
[7] J. Yim and K. Sohn, “Enhancing the performance of convolutional neural networks on quality degraded datasets,” CoRR, vol. abs/1710.06805, 2017.

**ADMISSION CRITERIA AND APPLICATION**

- Candidates must hold or be in the process of completing a master degree (or last year French engineering system degree) in telecommunications, signal/image processing, computer science, or applied mathematics (other equivalent academic degrees would be considered on a case-by-case basis).
- Programming tool: Python, Matlab or OpenCv.
- Good English skill (both writing and speaking).
- CV and scanned copies of diplomas and grades from all academic institutions of higher education should be sent in ***a single zipped*** file to Prof. Azeddine Beghdadi (azeddine.beghdadi@univ-paris13.fr).

**Supervisors**

Main supervisor: Prof. Azeddine Beghdadi (L2TI, USPN)

Co-supervisor : Dr Marie Luong (L2TI, USPN)

**Hosting institution**

The recruited candidate will be hosted at the Laboratoire de Traitement et Transport de l’Information (L2TI) at Université Sorbonne Paris Nord (USPN) in France.