Summary

Image and video processing play an important role in the development of technologies for dealing with security issues: surveillance cameras are widely diffused as means of crime reduction, and image analysis tools are used in the forensics field. In this thesis two problems are considered: the reconstruction of documents which have been reduced to a heap of paper strips by a shredder device and the enhancement of poorly illuminated surveillance videos.

The system architecture we developed for the computer-based re-assembly of shredded documents includes as a first step the acquisition of the strips with a scanner. After a pre-processing step, each strip is represented by a digital image. A binary mask is then generated, which permits to separate the strip from the acquisition background. In order to perform the reconstruction, the visual content of the strips must be properly coded, while the piece shape, commonly used for jigsaw puzzle or works of art fragments reconstruction, does not provide the necessary information. After a first attempt of describing the visual content by the MPEG-7 features, we resorted to domain-specific features. We recognized the following features as relevant for representing the strip visual content: line spacing, font type, number of lines of text, position of the first line of text, position of the last line of text, which are expedient when the original document contains printed text; squared paper index, useful in the case of notebook paper; presence of a marker, ink color, paper color, text edge energy, strip border, useful in both cases of handwritten and printed text. We developed the algorithms that automatically extract each one of those features from the strip digital image. The algorithms are specifically

designed for taking into account the shredded strips peculiarities.

On the base of the features, strips can be grouped in such a way that the strips belonging to the same page in the original documents are assigned to the same group and there are ideally as many groups as many the pages were. A hierarchical clustering algorithm has been used for this aim. The number of groups to be found is automatically selected by the algorithm in a proper interval provided by the user. The clustering is effective in improving the performance of a computer-aided reconstruction. Moreover, the computer computational time for the on-line interaction with the human operator is reduced by clustering. The computer-aided reconstruction is modelled as an image retrieval task: the user selects one strip, and the ones most similar to it are retrieved (ordered by decreasing similarity measure) and shown on the monitor. Among them, the user recognizes the correctly matching strips and virtually glues them. The process is repeated iteratively until the reconstruction has been accomplished.

In a fully automatic reconstruction scenario, the correctly matching strips have to be automatically detected by the computer. The information contained in the strip borders, along which the matching is performed, is exploited, namely the grey-level pixel appearance on the right (or left) strip border is used. The problem is modelled as a combinatorial optimization problem, and its NP-Completeness is demonstrated. Since it is NP-Complete, suboptimal algorithms must be devised for its solution. First, a local matching algorithm is proposed: given a piece, the correctly matching one on its right is the one whose left border is the most similar to the given strip right border. Errors may occur, since the border is noisy due both to the shredding and the digitalization processes. A global solution is thus explored, and the problem is modelled as an Assignment Problem: each left border must be assigned a right border, in such a way that the overall similarity is maximized.

In conclusion, the original contributions developed in this thesis concerning the shredded document reconstruction problem are the following:

1. the problem characterization;

2. the design of a number of numerical low level features describing the strip visual content; the features are automatically extracted by the computer;

3. the definition of an algorithm for grouping the strips belonging to a same page;
4. the modelling of the problem as a combinatorial optimization problem and the definition of polynomial sub-optimal algorithms for its automatic solution.

The second problem which has been studied during this PhD is the enhancement of poorly or non-uniformly illuminated images and videos. Both low dynamics and high dynamics images have been considered. In the latter, the enhancement is combined with a dynamics reduction, as explained below. High dynamics images are images that span a large range of luminosity. The Human Visual System has a high dynamics behavior: when looking towards a window from indoor we are able to distinguish both the internal and external details. Common acquisition devices lack this capability, and the resulting pictures could be saturated to white in the outdoor part or too dark in the indoor part. Techniques for the acquisition of high dynamics images exist. They consist in combining several pictures of the same scene taken with different exposure settings, or in using high dynamics sensors, such as the logarithmic CMOS sensor or the linear-logarithmic CMOS sensor. However, common display devices have a low dynamics and a dynamics reduction needs to be performed for visualization. The algorithm for dynamics reduction that has been considered in this thesis is the Locally Adaptive dynamics Reduction algorithm (LARx family). With respect to the existing literature, it has the advantage of being computationally light and thus suitable for real-time applications such as video surveillance and vehicle driving assistance. Like many other image enhancement algorithms, the LAR algorithm is based on the Retinex theory that states that when we observe an object, the image formed in our eye is the product of the illumination and of the object reflectance. It is the illumination that can present high dynamics, while the reflectance corresponds to the object details and has low dynamics. Therefore, for enhancing the images it suffices to compress the dynamics of the illumination, while keeping unchanged or enhancing the reflectance. The separation of the image into reflectance and illumination is however an ill-posed problem, and various solutions have been proposed in the literature. The LAR algorithms estimate the illumination using an edge-preserving smoothing filter. Their implementation by a Recursive Rational Filter results in a computationally light operator.

In surveillance, the problem to be solved is the acquisition under low luminosity. The LAR algorithm for video sequences (LARS) has been optimized for this application, where the number of algorithm parameters to be set by the non-expert user should be small, and a real-time processing may be necessary. Due to the fact that cameras are quite noisy at low luminosity, and the noise becomes even more visible after the enhancement process, we developed a different version of the algorithm to handle this problem.

In vehicle driving assistance, a high dynamics camera may be mounted on the vehicle rear mirror. The high dynamics sensor is particularly suitable for this application because the illumination could change very suddenly while the car is moving, for example when entering a tunnel or in case of direct sunlight. With the idea in mind of processing the videos directly on the camera, a low-cost hardware implementation of the LARS algorithm has been developed. It is tailored to FPGAs. Moreover, the temporal consistency has been taken into careful consideration to avoid annoying flickers.

Though many algorithms for dynamics reduction exist in the literature, the problem of objectively assessing their performance is still unsolved. Usually, a subjective qualitative evaluation is performed, strongly relating the algorithm performance to the personal observer taste. We developed two novel quality measure algorithms, namely the tool based on the co-occurrence matrix and the local contrast measure. Both take as a reference the high dynamics image and regard as good the algorithms whose output is a low dynamics image with similar characteristics. The co-occurrence matrix tool describes the spatial distribution of high and low dynamics images by means of a visual representation as well as of a number of numerical features. The second measure we developed focuses on the local contrast feature. Based on the local contrast, noise, details and homogeneous parts can be separated. An algorithm which is able in particular to enhance the details part is considered to perform well according to this measure. A third methodology for assessing the enhancement algorithm performance is to setup an experimental environment where the luminosity can be varied. The same scene can then be acquired with good luminosity (reference images) or under poor lighting conditions (images to be processed by the algorithms). In this case a good algorithm, given the badly illuminated images, would output images with a low distance from the reference image.

To summarize, the main original contributions of this thesis in the field of image and video enhancement are the following:

1. the LARS algorithm has been improved for surveillance and vehicle driving assistance applications (and an FPGA implementation has been proposed);

2. three novel objective quality measures to assess the performance of dynamics reduction algorithms have been developed.