Chapter 13
Distillation:
A Super-Resolution Approach for the Selective Analysis of Noisy and Unconstrained Video Sequences

Dong Seon Cheng
University of Verona, Italy

Marco Cristani
University of Verona, Italy

Vittorio Murino
University of Verona, Italy

ABSTRACT

Image super-resolution is one of the most appealing applications of image processing, capable of retrieving a high resolution image by fusing several registered low resolution images depicting an object of interest. However, employing super-resolution in video data is challenging: a video sequence generally contains a lot of scattered information regarding several objects of interest in cluttered scenes. Especially with hand-held cameras, the overall quality may be poor due to low resolution or unsteadiness. The objective of this chapter is to demonstrate why standard image super-resolution fails in video data, which are the problems that arise, and how we can overcome these problems. In our first contribution, we propose a novel Bayesian framework for super-resolution of persistent objects of interest in video sequences. We call this process Distillation. In the traditional formulation of the image super-resolution problem, the observed target is (1) always the same, (2) acquired using a camera making small movements, and (3) found in a number of low resolution images sufficient to recover high-frequency information. These assumptions are usually unsatisfied in real world video acquisitions and often beyond the control of the video operator. With Distillation, we aim to extend and to generalize the image super-resolution task, embedding it in a structured framework that accurately distills all the informative bits of an object of interest. In practice, the Distillation process: i) individuates, in a semi supervised way, a set of objects of interest, clustering the related video frames and registering them with respect to global rigid transformations; ii) for each one, produces a high resolution image, by weighting each pixel according to the
1. INTRODUCTION

In the emerging fields within video analysis, such as video indexing, video retrieval, video summarization and video surveillance, the quality of the video frames represents the essential source of information for the identification, classification or recognition of targets of interest, like objects or people. The widespread use of low-cost hand-held cameras, web-cams, and mobile phones with cameras has multiplied the sources of video production, but at the cost of a lower quality. The reasons are several: starting from careless acquisitions of inexperienced operators to low resolution cameras with low gains. The analysis of this data is problematic at best, and useless in many cases, because the noise and the resolution may not allow any meaningful processing, even if the object of interest is present in quite a few frames.

The problem of obtaining a highly informative image starting from noisy and coarsely resolved input images is known in literature as (image) super-resolution. When the input consists in only one low resolution image, we refer to the problem as “single-frame super-resolution” (Kursun & Favorov, 2002), when several frames are considered, the problem is called “multi-frame super-resolution” or simply super-resolution (Baker & Kanade, 2002; Schultz & Stevenson, 1994). There are other kinds of super-resolution currently under study, for example, the super-resolution enhancement of video (Bishop, Blake, & Marthi, 2003; Shechtman, Caspi, & Irani, 2002) whose goal consists in improving the quality of each single frame through the addition of high frequency information.

Recently, the attention devoted to the development of super-resolution algorithms is sensibly grown, in both the single image (Kim, Franz, & Scholkopf, 2004), and the multi-frame cases (Baker & Kanade, 2002; Ben-Ezra, Zhouchen, & Wilburn, 2007; Freeman, Jones, & Pasztor, 2002; Freeman, Pasztor, & Carmichael, 2000; Lin & Shum, 2004; Pickup, Capel, Roberts, & Zisserman, 2006; Pickup, Roberts, & Zisserman, 2006; Tippping & Bishop, 2002). Clearly, in the latter case, the information encoded in the resulting image is considerably larger, giving a more accurate representation (for an overview, see Sec. 2).

In video analysis, several tasks of recognition and detection are in fact based on visual data (Kanade, Collins, & Lipton, 2000). For example, in a video surveillance context, a common task consists in detecting the identity of a person captured with a camera, and therefore highly detailed images are desirable. In another context, video summarization considers the problem of generating a concise and expressive summary of a video sequence by extracting and abstracting the most relevant features in the scene. The higher the quality of this summary, the higher the capability of building an effective indexing capable of distinguishing among similar sequences. For these reasons, the
Related Content

FastNMF: Efficient NMF Algorithm for Reducing Feature Dimension
[www.igi-global.com/chapter/fastnmf-efficient-nmf-algorithm-reducing/43825?camid=4v1a](www.igi-global.com/chapter/fastnmf-efficient-nmf-algorithm-reducing/43825?camid=4v1a)

Human Motion Analysis
[www.igi-global.com/chapter/human-motion-analysis/5477?camid=4v1a](www.igi-global.com/chapter/human-motion-analysis/5477?camid=4v1a)

Biometric Image Processing
[www.igi-global.com/chapter/biometric-image-processing/76160?camid=4v1a](www.igi-global.com/chapter/biometric-image-processing/76160?camid=4v1a)

A Cyber-Physical Photovoltaic Array Monitoring and Control System
[www.igi-global.com/article/a-cyber-physical-photovoltaic-array-monitoring-and-control-system/205543?camid=4v1a](www.igi-global.com/article/a-cyber-physical-photovoltaic-array-monitoring-and-control-system/205543?camid=4v1a)