CCIT REPORT #756 February 2010

A Hybrid Vector Wiener Filter Approach to Translational Super-Resolution

Tomer Michaeli, *Student Member, IEEE* and Yonina C. Eldar, *Senior Member, IEEE*

***Abstract*—We address the problem of purely-translational super-resolution (SR) for signals in arbitrary dimensions. We show that discretization, a key step in many SR algorithms, inevitably leads to inaccurate modeling. Instead, we treat the problem entirely in the continuous domain by modeling the signal as a continuous-space random process and deriving its linear minimum mean-squared error (LMMSE) estimate given the low-resolution discrete-space observations. We derive a closed form expression for the resulting mean-squared error and use it to analyze the emergence of periodic artifacts in the super-resolved signal. We also provide three efficient implementation schemes of the LMMSE estimate, one of which specialized for 1D applications. These methods constitute a natural generalization of several well known single-image recovery algorithms, such as spline interpolation, to the multichannel SR setting. Experiments on real-world images demonstrate the advantage of our approach with respect to several prominent SR techniques that rely on discretization.**

***Index Terms*—Super-resolution, nonuniform interpolation, hybrid Wiener filter.**

I. Introduction

SUPER-RESOLUTION refers to the process of combining several low-resolution descriptions of a signal to form one higher resolution version of it. In the field of image processing, there are 1D, 2D and 3D variants of this task. Perhaps the most commonly treated scenario is that of purely-translational spatial (2D) super-resolution (SR) [1], [2]. Here, several low-resolution noisy images of a scene are captured by a camera, each with a different translation, and the goal is to produce one high-resolution image of the same scene. Three-dimensional scenarios arise in space-time SR applications [3]. There, several video sequences of the same scene are fused into one higher-resolution video stream at a higher frame-rate. If the video cameras are spatially co-calibrated, then processing can be carried out only along the time dimension (temporal SR), rendering the problem one-dimensional. In this paper, we collectively refer to all problems of this type as SR, and develop a theory for arbitrary d-dimensional signals.

The physical model underlying purely-translational SR scenarios can be described mathematically as [4]

$$c_k[\boldsymbol{n}] = [(s * x)(\boldsymbol{t})]_{\boldsymbol{t}=\boldsymbol{n}-\boldsymbol{t}_k} + u_k[\boldsymbol{n}], \qquad k = 1,\dots,K, \quad (1)$$

This work was supported in part by the Israel Science Foundation under Grant no. 1081/07 and by the European Commission in the framework of the FP7 Network of Excellence in Wireless COMmunications NEWCOM++ (contract no. 216715).

The authors are with the department of electrical engineering, Technion–Israel Institute of Technology, Haifa, Israel (phone: +972-4-8294682, fax: +972-4-8295757, e-mail: tomermic@tx.technion.ac.il, yonina@ee.technion.ac.il).

where $c_k[\boldsymbol{n}]$, $\boldsymbol{n} \in \mathbb{Z}^d$, is the kth discrete-space observation of the continuous-space signal $x(\boldsymbol{t})$, $\boldsymbol{t} \in \mathbb{R}^d$, acquired with translation $\boldsymbol{t}_k$ and additive noise $u_k[\boldsymbol{n}]$. The filter $s(\boldsymbol{t})$ is associated with the imaging device. It corresponds to the point-spread-function (PSF) of the lens in 2D SR, to the temporal integration profile of the sensor in 1D SR, and to a combination of both in 3D space-time SR. The goal of an SR algorithm is to produce a discrete-space high-resolution signal $\boldsymbol{x}_{\text{HR}}[\boldsymbol{n}]$ which corresponds to the samples of $x(\boldsymbol{t})$ on a dense grid.

In much of the recent literature, SR is modeled via the discrete-space relations [1], [2]:

$$\boldsymbol{c}_k = \boldsymbol{S}_k \boldsymbol{x}_{\text{HR}} + \boldsymbol{u}_k, \qquad k = 1,\dots,K. \quad (2)$$

Here $\boldsymbol{c}_k$, $\boldsymbol{u}_k$, and $\boldsymbol{x}_{\text{HR}}$ are column vectors comprising the elements of $c_k[\boldsymbol{n}]$, $u_k[\boldsymbol{n}]$, and $\boldsymbol{x}_{\text{HR}}[\boldsymbol{n}]$ respectively, and the matrix $\boldsymbol{S}_k$ accounts for the filtering by $s(\boldsymbol{t})$ and the sampling. This discrete formulation has several advantages over the continuous model (1). Most noteworthy, it allows the construction of a finite-dimensional optimization problem incorporating complicated realistic prior-knowledge assumptions on the unknown $\boldsymbol{x}_{\text{HR}}$ [2]. Such optimization problems may be solved using standard optimization methods. Nevertheless, as we show in this paper, there are many situations in which the continuous-space equations (1) cannot be represented in the discrete form (2). In other words, we show that in certain settings SR must be regarded as a continuous-space interpolation problem. Algorithms which do not treat the problem from this viewpoint, inevitably suffer from a model miss-match error.

Prior SR work falling into the interpolation category (also termed frequency-domain methods [4]) include [5], [6], [7], [8], [9], [10], [11], [12]. In these algorithms, the continuous-space scene $x(\boldsymbol{t})$ is assumed to be bandlimited. In the single-channel interpolation literature, recent work has shown that this type of prior knowledge is often not in agreement with the typical behavior of natural images [13], [14]. Furthermore, it has been experimentally demonstrated that interpolation algorithms relying on non-bandlimited models, such as the Matèrn prior, can lead to improved reconstruction results [15], [16], [17]. In SR applications, the bandlimited assumption translates to a stringent limitation on the required number of measurements. Specifically, the number of observations K required for increasing the resolution by a factor of Δ must be at least Δ^d in the methods mentioned above. Thus, to increase resolution by a factor of 4 in every dimension, for example, at least 16 still images are required in a spatial SR task ($d = 2$), and at least 64 video sequences are needed in a space-time SR scenario ($d = 3$).