J. Vis. Commun. Image R. 40 (2016) 98–110

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

A robust iterative super-resolution mosaicking algorithm using an adaptive and directional Huber-Markov regularization ☆

Debabrata Ghosh [a,*], Naima Kaabouch [a], Wen-Chen Hu [b]

[a] Department of Electrical Engineering, University of North Dakota, Grand Forks, ND, USA
[b] Department of Computer Science, University of North Dakota, Grand Forks, ND, USA

ARTICLE INFO

Article history:
Received 4 May 2015
Revised 17 April 2016
Accepted 14 June 2016
Available online 18 June 2016

Keywords:
Mosaicking
Super-resolution
Scale invariant feature transform (SIFT)
Homography
Steepest descent optimization
Huber-Markov regularization

ABSTRACT

A robust spatial-domain based super-resolution mosaicking algorithm is proposed. This technique incorporates a mosaicking algorithm, and a super-resolution reconstruction algorithm. The main contribution of this paper is the development of a super-resolution algorithm using a Huber Norm-based maximum likelihood (ML) estimation in combination with an adaptive directional Huber-Markov regularization. Another contribution is the development of a no-reference performance metric based on reciprocal singular value curve for quantitative evaluation of the proposed algorithm. Along with the above-mentioned metric, five other performance measurement metrics are used to assess the efficiency of the algorithm. The performance of this algorithm is compared with the performances of two different algorithms: the Tikhonov regularization-based and the total variation (TV)-based super-resolution mosaicking algorithms. Results show that the proposed algorithm outperforms the other two techniques in terms of lowest amount of blur and noise in the output.

1. Introduction

High-resolution images are desired in many practical applications, including satellite imaging, medical image processing, target recognition, video surveillance, etc. The immediate solution to achieve high-resolution image is to increase the detector density by reducing their size or to increase the sensor area. However, these involve manufacturing challenges and is prohibitively expensive in certain applications. Consequently, the use of signal processing techniques to obtain an image with improved spatial resolution from multiple low-resolution (LR) images of the same scene becomes an attractive proposition. This approach, known as super-resolution (SR) reconstruction, is well documented in the literature. Such reconstruction primarily relies on the ability to estimate the relative displacement between the sequential low-resolution frames to recover details that are finer than the sampling grid [1]. Additionally, the possible effects causing blur and noise in the low-resolution frames are eliminated in this reconstruction procedure. One obvious advantage of this approach is that existing low-resolution imaging systems can still be utilized [2]. Image mosaicking, in similar fashion, is the stitching of two or more correlated images of the same scene to yield an integral representation of the overall scene [3]. When image mosaicking and super-resolution are combined together, multiple overlapping low-resolution images of a scene can be fused together to obtain high-resolution panoramic view of the scene [4]. This method is referred to as super-resolution mosaicking.

Over the last decade, a number of algorithms related to mosaicking [5–15] and super-resolution reconstruction [16–27] have been proposed. Super-resolution reconstruction performance has been most influenced by the choice of ML estimation and regularization technique. An L2 Norm-based ML estimation with Markov Random Fields regularization is proposed in [16]. Fasiu et al. [17] proposed an L1 Norm-based ML estimator and bilateral total variation in their super-resolution algorithm. Later they [18] used an L1 Norm-based ML estimator and a combination of luminance, chrominance, and orientation regularizations for color image super-resolution. Zhang et al. [19] used L2 Norm-based ML estimator with Tikhonov regularization for super resolving MRI images. Authors in [20] employed a Huber Norm-based ML estimator along with Huber-Tikhonov regularization. Ng et al. [21] suggested an L2 Norm-based ML estimator together with TV regularization. An L2 Norm-based estimator and Laplacian regularization is proposed in [22]. Panagiotopoulou et al. [23] utilized Turkey, Lorentzian, and Huber Norm-based ML estimators in combination with bilateral total variation regularization. In [24], authors proposed an L2

☆ This paper has been recommended for acceptance by Zicheng Liu.

* Corresponding author.
E-mail address: debabrata.ghosh@ndus.edu (D. Ghosh).

http://dx.doi.org/10.1016/j.jvcir.2016.06.008
1047-3203/

Norm-based ML estimator along with directional bilateral total variation and tri-modal regularization. L2 Norm-based ML estimators with spatially adaptive total variation regularizations are described in [25,26]. Chen et al. [27] proposed an L2 Norm-based ML estimator and adaptive co-sparse regularization. Most of the aforementioned methods [16–18,21,23,25,26] are based on simple estimation techniques, such as L1 and L2 Norm, for the data fidelity term; hence, they are very sensitive to their assumed model of data and noise. In robust statistical estimation problems, Huber Norm is designed to manage outliers more efficiently than the L1 and L2 Norms. Even though a few of the aforementioned methods [19,22] investigate the effect of Huber Norm as a superior estimator, they are either computationally expensive because of the choice of regularization or otherwise inferior to the methods, which use simpler regularizations. Majority of the aforementioned methods [16–23] used constant regularizations with manually chosen regularization parameters, which have the disadvantage of generating artifacts particularly in the smooth image regions. An adaptive regularization, on the other hand, uses spatial information distributed in the image to constrain the regularization strength in different image regions. Even though [25–27] used adaptive regularization models, use of less robust L2 Norm and computationally expensive regularization model make these methods unsuitable for many applications. In [28–30], authors used adaptive regularization parameters. However, these methods are based on less suitable L2 Norm and Tikhonov regularization, which usually generates overly smooth output. Moreover, from the state of the art, it is obvious that even though research has been performed in mosaicking and super-resolution, very little work has been done on super-resolution mosaicking.

To assess the efficiency of the super-resolution algorithms, several metrics [20,25–27,31–35] have been proposed over the last decades. Ng et al. [20] used a single performance metric requiring a ground truth image to evaluate their algorithm. Yuan et al. [25] used two no-reference performance metrics. However, these metrics do not seem to be correlated to the human visual perception, and, hence fail when tested for different distortion types or different distortion levels within a particular distortion type. Authors in [26,27] used two metrics, which rely on the availability of the ground truth frames. Zhang et al. [31] used three metrics that require ground truth images and suffer from indistinguishable dynamic ranges. Laflen et al. [32] proposed four metrics; however, all these metrics require computationally expensive setups. Tian et al. [33] suggested three measurement metrics that involve either measuring sub-pixel shifts between frames or detecting interest points or segmenting an image, making these metrics computationally complex. Additionally, all the metrics require ground truth images for evaluation. Yet another three performance metrics used in [34] require the availability of ground truth frames. Additionally, these metrics do not seem to be superior to simple pixel-based measures. Nelson et al. [35] suggested the use of two performance metrics, which suffer either from having poor dynamic range, or from quantifying fidelity rather than the amount of distortion in the output.

In this paper, we propose a super-resolution mosaicking algorithm, which registers the successive frames into a common coordinate system and simultaneously generates mosaic output with improved resolution. The novelty of the proposed method is that it is based on an adaptive regularization and it utilizes Huber Norm for maximum likelihood estimation in combination with a directional Huber-Markov regularization. Use of a no-reference performance metric based on reciprocal singular value curve for evaluating the proposed algorithm is also a novel idea. Other than this metric, five performance measuring metrics mentioned in [4] are also used for quantitative evaluation. In order to demonstrate the superiority of the proposed method, its performance is compared with two other types of super-resolution mosaicking methods based on Tikhonov regularization [19], and TV regularization [21] for super-resolution. The remainder of this paper is organized in five sections. In Section 2, we discuss the mathematical model of the proposed super-resolution mosaicking approach and the three other aforementioned approaches used for performance comparison. In Section 3, we present our proposed algorithm, performance evaluation metrics, and experimental setup. In Section 4, we provide and discuss the evaluation results, demonstrating the efficacy of the proposed algorithm. Finally, in Section 5, we draw evidence-based conclusions.

2. Methodology

2.1. Mathematical model

The proposed super-resolution mosaicking method and the two other comparative methods are all based on similar concepts of minimizing an error functional using maximum *a posterior* estimates and then solving optimization problems. Thus, these algorithms share similar mathematical backgrounds but utilize different Norms and regularizations. In this section, the common mathematical model using various Norms and regularizations employed by these three algorithms is discussed in detail.

In order to develop a comprehensive understanding of the super-resolution mosaicking algorithm it is often customary to formulate a linear observation model, which relates the acquired low-resolution images to the super-resolution mosaic. The observation model aims to include most of the factors that cause degradations to the acquired images. The current model incorporates warp, blur (both atmospheric blur and optical blur), noise, and downsampling, since these are the most common degradations and can be modeled fully or partially in different super-resolution mosaicking techniques. According to [36] the observation model could be expressed as:

$$y_k = \mathrm{D}B_k W_k B_k^a R[x]_k + n_k \quad \text{for } 1 \leqslant k \leqslant K \tag{1}$$

where the kth low-resolution observation y_k is generated from the desired super-resolution mosaic x, which undergoes the aforementioned degradations. B_k^a denotes the atmospheric blur effect, W_k denoted the warp operation, B_k represents the optical blur effect, and D is the decimation effect. n_k is the additive noise for the kth image. $R[R[\cdot]]$ is the reconstruction operator, that extracts warped images from the super-resolution mosaic. Conventionally, the variables $R[x]_k$, y_k, and n_k are rearranged as column vectors in lexicographic order, whereas the variables B_k^a, W_k, B_k, and D are expressed as matrices.

Since the aim of the super-resolution mosaicking algorithm is to determine an estimate of x given the captured image sequence and the characterization of the imaging process, it is essentially an inverse process. Consequently the super-resolution mosaicking algorithm's stability is not solely determined by the availability of multiple low-resolution observations, rather estimation of several other factors like B_k^a, B_k, and n_k are also necessary [4]. Clearly, super-resolution mosaic assembly is a large sparse optimization problem, which could be solved using iterative methods [37]. However, instead of sparse matrices multiplication, basic image operations (e.g. convolution, warping, down-sampling) could be applied along with gradient computation in order to speed up the required super-resolution computations. Subsequently, an estimate of the super-resolution mosaic $\hat{x}$ could be achieved from Eq. (1) by optimizing a utility function, which minimizes the error between the input low-resolution images and the reconstructed ones [38]. A common utility function using the maximum likelihood estimate is expressed as: