J. Vis. Commun. Image R. 25 (2014) 855–863

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Bandlet-based sparsity regularization in video inpainting

Ali Mosleh [a], Nizar Bouguila [b,*], A. Ben Hamza [b]

[a] Department of Electrical and Computer Engineering, Concordia University, Montréal, QC H3G 2W1, Canada
[b] Concordia Institute for Information Systems Engineering, Concordia University, Montréal, QC H3G 2W1, Canada

ARTICLE INFO

Article history:
Received 6 July 2012
Accepted 13 January 2014
Available online 24 January 2014

Keywords:
Bandlets
Inpainting
Patch fusion
Regularization
Video completion
Spatio-temporal flows
Video sequence
Missing information

ABSTRACT

We present a bandlet-based framework for video inpainting in order to complete missing parts of a video sequence. The framework applies spatio-temporal geometric flows extracted by bandlets to reconstruct the missing data. First, a priority-based exemplar scheme enhanced by a bandlet-based patch fusion generates a preliminary inpainting result. Then, the inpainting task is completed by a 3D volume regularization algorithm which takes advantage of bandlet bases in exploiting the anisotropic regularities. The method does not need extra processes in order to satisfy visual consistency. The experimental results demonstrate the effectiveness of our proposed video completion technique.

1. Introduction

Missing parts in still images and video sequences may be caused by damages or deliberately undesired object removal from the images or the video frames. The image/video inpainting problem has attracted a great attention in the past few years due to its powerful ability in fixing and restoring damaged saptial/spatio-temporal data. In this paper, we focus on video inpainting as a technique to recover missing data in some specified regions of videos. Due to the large dimensionality of video data coupled with its saptio-temporal consistency which must be preserved, video inpainting can be considered as a challenging task even though large amount of data can be highly desirable to fill-in the missing regions.

One can refer to [1] for detailed mathematical interpolation models specialized in image inpainting. The pioneering work in digital inpainting [2] employs non-linear partial differential equations (PDEs) as an interpolation platform to perform image and video frame inpainting. The concepts of PDEs and interpolation in inpainting have been employed in many techniques, including [3] which derives a third-order PDE based on Taylor expansion to propagate the border isophotes to the missing regions. An explicit extension of the technique introduced in [2] is presented in [4] which applies Navier-Stoke equations. This approach applies ideas from classical fluid dynamics to continuously propagate isophote lines of the image from the exterior into the inpainting zone. As another technique, the proposed video inpainting scheme in [5] benefits from discrete *p-Laplacian* regularization on a weighted graph. Despite their promising results, the PDE and interpolation based methods perform frame-by-frame completion that neglects the continuity across consecutive frames unless PDEs are adapted in a 3D scheme [6]. Moreover, these methods are appropriate only for narrow and small missing regions.

The concept of priority in image inpainting introduced in [7] has been adopted in various video inpainting approaches. In these techniques, a correct order of filling-in process leads to a high performance in the completion task. Important properties, such as availability, trackability and motion vectors of the pixels, and geometric properties contribute to the calculation of the priority of the missing regions to be filled-in first. For instance, the method introduced in [8] performs moving object segmentation to separate the background and foreground of the video. Hence, the search space is reduced for completion of partially occluded moving objects [9]. In this method a motion confidence value is used to find the priority of the filling-in area in order to maintain the temporal consistency in the foreground completion task. For the background completion step, the image inpainting technique introduced in [7] is adopted. Modifications based on analysis of continuities on stationary and non-stationary videos are carried out to find the best priority in [10]. Then, in [11] the technique is further improved for various

* Corresponding author. Permanent address: Office EV-007-632, Concordia Institute for Information Systems Engineering, Concordia University, 1515 St. Catherine Street West, EV.007.632, H3G 2W1 Montreal, Canada.

E-mail addresses: mos_ali@encs.concordia.ca (A. Mosleh), bouguila@ciise.concordia.ca (N. Bouguila), hamza@ciise.concordia.ca (A.B. Hamza).

http://dx.doi.org/10.1016/j.jvcir.2014.01.007
1047-3203/

camera motions by keeping the track of similar regions. The priority is determined based on the trackability of the pixels in the method introduced in [12]. The highest priority fragment around the boundary of the missing region is completed using a graph-cut fragment updating instead of copying just a similar texture from the undamaged region. In [13] a priority-based method considers the video completion task as a global search optimization in order to find the best match. The whole video is considered as a volume and a multi-scale scheme is employed to reduce the computation time. Motion layer segmentation is the key step in the method proposed in [14]. Each separate layer is completed using the image inpainting method, and then all the layers are combined in order to restore the final video. A two phase sampling and alignment video inpainting technique is introduced in [15]. The method predicts motion data in the foreground, then missing moving foreground pixels are reconstructed by spatio-temporal alignment of the sampled data. Then, the background inpainting is done by 3D tensor voting as an extension of the still image repairing technique introduced in [16]. The methods in [17,18] proposed to inpaint videos by transferring sampled motion fields from the available parts of the video. The latter method tracks patches containing missing regions in the adjacent frames by employing a global motion estimation scheme. In [19,20] 3D patch-based probability models with potential applications in video inpainting are introduced. The probability model introduced in [19] is an alternative for motion models such as optical-flow. A sparsity-based prior for a variational Bayesian model is defined for video sequences. The damaged portion of a video can be treated in this Bayesian framework as an inpainting task. A learning strategy in [20] on the video's 3D space–time patches leads to video epitomes. Epitomes are viewed as a set of 3D arrays of probability distributions applied for video reconstruction. Although the preliminary results of video inpainting using these methods are promising, they need more improvements to be able to deal with large missing portions.

Maintaining the visual consistency along with handling the large dimensionality of videos in the inpainting process is an important fact. No wonder we see complicated steps in the state-of-the-art techniques, such as segmentation of different motion layers or objects, foreground/background separation, tracking, optical-flow mosaics computation and so onto cope with spatio-temporal consistency. In this paper we propose an approach that takes advantage of the bandlets sparse representation to reconstruct missing data visually pleasingly. Image sparse representation methods were introduced for spatial inpainting problems [21–23]. In such methods, missing pixels are inferred by adaptively updating the sparse representation (e.g. wavelets, DCT, etc). Although these approaches are very challenging to be adapted to video completion that deals with unsound and damaged estimated motion vectors, they yield satisfactory results in the case of image inpainting. Apparently, employing an efficient sparse representation can enhance the inpainting results. The main motivation behind employing the bandlet domain is due to its effective capability in capturing the geometric properties of an image as an efficient sparse representation [24]. The captured geometric features are used in our technique to firstly blend the results of patch matching in order to keep the visual consistency. Secondly, the overall bandlet geometry of the frames can be a good prior if we consider the video inpainting as an ill-posed linear problem. The obtained overall geometry is used for sparse regularization to reconstruct the video. In our method, making distinction between static camera videos and sequences containing camera motions is not needed. Besides, there is no segmentation, tracking or complex motion estimation as applied in many of the previously discussed methods to facilitate the inpainting process. This is the main difference with our previous work [25] that relies on an accurate background/foreground segmentation in order to treat videos captured by static and moving cameras in different fashions by patch matching rather than bandlet-based patch fusion and 3D regularization.

The rest of this paper is organized as follows. Section 2 describes the idea behind the bandlet transform capability in reconstructing missing regions. Then, the proposed bandlet-based video inpainting method is presented in Section 3. In Section 4, the experimental results are provided. Finally, Section 5 concludes this paper.

2. Using bandlets in inpainting

The bandlet framework can achieve an effective geometric representation of texture images. It is essential in sparse regularization and spatial or spatio-temporal data reconstruction for digital inpainting purposes.

Although geometric regularity along image edges is an anisotropic regularity, conventional wavelet bases can only exploit the isotropic regularity on square domains. An image can be differentiable in the direction of the tangent of an edge curve even though the image may be discontinuous across the curve. Bandlet transform [26] exploits such anisotropic regularity. Bandlet bases construct orthogonal vectors elongated in the direction of the maximum regularity of a function. The earlier bandlet bases [27,28] have been improved by a multi-scale geometry defined over wavelet coefficients [29,30]. Indeed, bandlets are anisotropic wavelets warped along the geometric flow.

Considering the Alpert transform as a polynomial wavelet transform adapted to an irregular sampling grid, one can obtain vectors that have vanishing moments on this irregular sampling grid. This is the principal need to approximate warped wavelet coefficients. Only a few vectors of Alpert basis can efficiently approximate a vector corresponding to a function with anisotropic regularity. This *bandletization* using wavelet coefficients is defined as

$$b^k_{j,l,n}(x) = \sum_p a_{l,n}[p]\psi^k_{j,p}(x), \tag{1}$$

where j and k represent wavelet scale and orientation, respectively. The $a_{l,n}[p]$ are the coefficients of the Alpert transform where l is the scale and n is the index of the Alpert vector. In essence, $a_{l,n}[p]$ are the coordinates of the bandlet function $b^k_{j,l,n}$. These coefficients strictly depend on the local geometric flow. Bandlet coefficient are generated by inner products $\left\langle f, b^k_{j,l,n}\right\rangle$ of the image f with the bandlet functions $b^k_{j,l,n}$. The set of wavelet coefficients are segmented in squares S for polynomial flow approximation of the geometry. For each scale 2^j and orientation k, the segmentation is carried out using a recursive subdivision in dyadic squares. A square S should be further subdivided into four sub-squares, if there is still a geometric directional regularity in the square. Apparently, only for the edge squares, the adaptive flow is needed to be computed to obtain the bandlet bases. The geometry of an image evolves through scales. Therefore, for each scale 2^j of the orientation k a different geometry Γ^k_j is chosen. The set of all geometries $\{\Gamma^k_j\}$ represents the overall geometry of an image. Each member of this set is in fact a geometry value associated to one segmentation square S. For details about bandlets the reader is referred to [26].

The image inpainting problem may be formulated as follows. An image I contains a set of missing pixels indicated by Ω and a source ($\Phi = I \setminus \Omega$) area. The goal is finding an image $\acute{I}$ such that $\acute{I}(x)$ is equal to $I(x)$ for the pixels that belong to Φ, i.e., $\acute{I}(x) = I(x)\ \forall x \notin \Omega$ while the overall geometry of $\acute{I}$ has the same geometrical regularity as that of I in Φ. In the presence of additive noise ω we have the image f with missing pixels as $f = \theta I + \omega$ where

$$\theta I(x) = \begin{cases} 0 & \text{if } x \in \Omega \\ I(x) & \text{if } x \in \Phi. \end{cases} \tag{2}$$