A Rank Minimization Approach to Video Inpainting

Tao Ding
Dept. of Electrical Engineering
The Pennsylvania State University
University Park, PA 16802

Mario Sznaier and Octavia I. Camps
Dept. of Electrical and Computer Engineering
Northeastern University
Boston, MA 02115

Abstract

This paper addresses the problem of video inpainting, that is seamlessly reconstructing missing portions in a set of video frames. We propose to solve this problem proceeding as follows: (i) finding a set of descriptors that encapsulate the information necessary to reconstruct a frame, (ii) finding an optimal estimate of the value of these descriptors for the missing/corrupted frames, and (iii) using the estimated values to reconstruct the frames. The main result of the paper shows that the optimal descriptor estimates can be efficiently obtained by minimizing the rank of a matrix directly constructed from the available data, leading to a simple, computationally attractive, dynamic inpainting algorithm that optimizes the use of spatio/temporal information. Moreover, contrary to most currently available techniques, the method can handle non–periodic target motions, non–stationary backgrounds and moving cameras. These results are illustrated with several examples, including reconstructing dynamic textures and object disocclusion in cases involving both moving targets and camera.

1. Introduction

The problems of still–image and video inpainting, that is the process of automatically and seamlessly restoring or altering portions of an image or a video clip has been the subject of considerable attention in the past few years (see for instance [17, 30, 22, 10, 16, 37, 23, 19, 33]).

Earlier work in the field led to a number of algorithms that worked well for still images inpainting and completion in various cases ([2], [3], [9], [20], [32], [26], [12]), but had very limited applicability in the case of image inpainting in video sequences. Several image–sequence based approaches have been proposed to handle these limitations.

Kokaram *et al*. [17] proposed the use of three interpolators to perform motion estimation and interpolate missing portions in films from frames around. This technique works well for restoring relatively small losses not spanning continuous frames, but has a substantial computational burden. An optimal-flow estimation based algorithm was proposed to repair poor-quality video frames in [16], but its applicability is restricted to slightly corrupted and slowly changing image areas. Wexler *et al*. [37] minimize an objective function to find a patch satisfying spatio-temporal consistency constraints with the patches around the missing area. However, at the present time these results are restricted to objects undergoing repetitive motion, and the computational burden of the algorithm is high, due to its iterative nature. A probabilistic technique has been proposed in [6], based upon learning a patch based probability model, which in turn is used to synthesize the missing area. A drawback of this approach is that it does not guarantee smoothness of the resulting clip.

Patwardhan *et al*. [23] presented an algorithm for filling-in partial occlusion in video sequences obtained from a fixed camera, and have recently extended these results to moving cameras. A similar approach has been also pursued in [15], where tracking is used to speed up the search for the target. In [14], Jia proposed a two-phase method for video repairing dealing with various constraints imposed by periodic motion, camera subclass, etc. In [39], a motion layer segmentation method was introduced to remove relatively large objects from the original video. Additional video completion papers based on figure-background segmentation include ([7], [27], [28], [4], [29]). However, at the present time, none of these methods exploits the underlying global spatio-temporal dynamics of the sequence.

In this paper we propose a new approach to video inpainting that exploits some simple results from Linear Systems theory to recast the problem into a matrix rank minimization problem. To this effect, we propose to proceed as follows: (i) find a set of descriptors that encapsulate the information necessary to reconstruct a frame, (ii) find an optimal estimate of the value of these descriptors for the missing/corrupted frames, and (iii) use the estimated values to reconstruct the frames. The main result of the paper shows that the optimal descriptor estimates can be efficiently obtained by minimizing the rank of a matrix directly constructed from the available data, leading to a simple,

978-1-4244-1631-8/07/$25.00 ©2007 IEEE

computationally attractive, dynamic inpainting algorithm. The proposed method has the following advantages over currently existing techniques:

1. It leads to a non–iterative, computationally attractive algorithm that optimizes the use of (global) spatio/temporal and dynamics information and has a moderate computational burden.

2. It is not restricted to the case of periodic motion, static background or stationary cameras.

3. It can be used to extrapolate frames, that is extend a given video sequence, and, in the case of dynamic textures, to artificially generate other textures from the same family.

These results are illustrated with several examples that include reconstructing missing portions of dynamic textures and object disocclusion in cases involving a moving target and camera.

2. Rank minimization approach to inpainting

In this section we outline the basic ideas that allow for recasting the inpainting problem into a rank–minimization one. For the sake of clarity, the technical details are relegated to the Appendix.

2.1. Spatio–temporal descriptor modelling

As indicated before, the main idea of the proposed approach is to, rather than directly attempt to interpolate missing pixels, estimate, based on all available spatio–temporal information, the value of a set of descriptors that encapsulate the information necessary to reconstruct missing/corrupted frames. The specific definition of these descriptors is problem dependent and will be discussed in Section 3.

In order to estimate the values of missing descriptors, we will collect the values of all the descriptors corresponding to the k^{th} frame in a vector $\mathbf{f}_k$ and assume that these values are generated by a stationary Gauss–Markov random process. This is equivalent to assuming that $\mathbf{f}_k$ is related to its values in previous frames by an ARMAX model of the form:

$$\mathbf{f}_{k+1} = \sum_{i=0}^{m-1} g_i \mathbf{f}_{k-i} + h_i \mathbf{e}_{k-i} \tag{1}$$

where g_i, h_i are fixed coefficients and $\mathbf{e}(.)$ denotes a stochastic input. Note that this can be always assumed without loss of generality, since given N_F measurements of $\mathbf{f}(.), \mathbf{e}(.)$, there always exists a linear operator such that (1) is satisfied ([24], Chapter 10). Finally, by absorbing if necessary the spectral density of $\mathbf{e}$ in the coefficients g_i and h_i, it can always be assumed that $\mathbf{e}(.)$ is an impulse.

It is worth emphasizing that we are not simply expressing each pixel as a linear combination of surrounding pixels lagged both in space and in time, as in [34]. Rather, the model (1) represents the present value of a descriptor (which may be related in a nonlinear way to the values of the pixels) as a linear combination of its past values. In this context, if the coefficients h_i and g_i of the model and the inverse mapping from descriptors to pixels are known, then the inpainting problem can be trivially solved by using the model (1), along with the available measurements of $\mathbf{f}$, to reconstruct the missing data. In principle one could try to use a two–tiered approach, where the model that best explains the available data is first extracted from the uncorrupted frames and then used to inpaint the missing values. However, as we show next, finding an explicit model is unnecessary: missing values of each descriptor in $\mathbf{f}$ can be *directly* found by solving a rank–minimization problem, obviating the need for finding an explicit model. As illustrated with several examples, this observation leads to simple, computationally tractable inpainting algorithms.

2.2. Completing Descriptor Sequences via Rank Minimization

The main idea behind the proposed approach is to find the set of values of the missing descriptors that *maximizes* the smoothness of the resulting inpainted sequence. Equivalently, denoting by f_k^o the observed descriptors and by f_k^m the missing ones, the idea is to find the values of f^m that are maximally consistent with f^o, in the sense that the resulting inpainted sequence is described by the *simplest* (e.g. most compact) possible model of the form (1). Specifically, as briefly described in the Appendix, the minimum value of m such that the model (1) explains the observed data is given by the rank of a matrix constructed from the measurements. Thus, the simplest model that explains this data corresponds to the values of the missing pixels that minimize this rank. Since it is well known that rank minimization problems are NP hard, rather than minimizing rank, we will relax this problem to a convex semi-definite programming problem, with the additional advantage of improving robustness against noise. These ideas lead to the following Algorithm:

Algorithm 1

1.- Given the observed values of the descriptors f^o, form the following (Hankel) matrix:

$$H_f \doteq \begin{bmatrix} f_1 & f_2 & \cdots & f_{n/2} \\ f_2 & f_3 & \cdots & f_{n/2+1} \\ \vdots & \vdots & \ddots & \vdots \\ f_{n/2} & f_{n/2+1} & \cdots & f_{n-1} \end{bmatrix} \tag{2}$$

Here f denotes either the observed data f_k^o, if the k frame is present, or the unknown value f_k^m, if the frame *needs to be inpainted, and n denotes the total number of frames.*

2.- *Estimate the values f^m which are maximally consistent with f^o by solving the following Linear Matrix Inequality (LMI) optimization problem,*

$$\begin{array}{ll} \textit{minimize w.r.t } f^m & Tr(Y) + Tr(Z) \\ \textit{subject to} & \begin{bmatrix} Y & H_f \\ (H_f)^T & Z \end{bmatrix} \geq 0 \end{array}$$

where $Y^T = Y \in \mathcal{R}^{n \times n}$, $Z^T = Z \in \mathcal{R}^{n \times n}$ and $H_f \in \mathcal{R}^{n \times n}$.

Note that this optimization problem can be efficiently solved using both commercially and freely available software. For the sake of clarity, the proof that the above algorithm leads to the simplest model explaining the observed data, and an explanation of the rationality for choosing this criteria for inpainting are relegated to the Appendix.

3. Application: Video inpainting

In this section we address the problem of filling-in damaged/missing parts of video clips. The approach that we propose to video inpainting is loosely related to the one proposed in [37] and [23], in the sense that we also use information from unoccluded frames to complete the missing information. However, rather than searching for the "best" matching patch, we use the interpolated/extrapolated values of the descriptors to reconstruct the missing information using a (nonlinear) combination of these pixels. The advantages of this approach are twofold: (i) it does not require that portions of the missing patch be (approximately) present in the unoccluded frames, effectively removing the constraint of (quasi) periodic motion in severe occlusion cases; and (ii) results in substantial computational complexity reduction vis-a-vis direct pixel reconstruction.

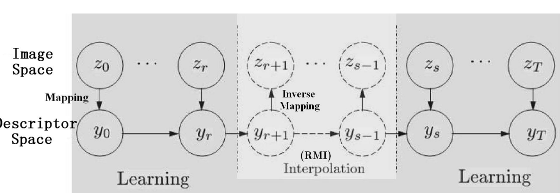


Figure 1. A general model for video inpainting by nonlinear dimensionality reduction, rank minimization and nonlinear reconstruction. z_t and y_t represent the appearance in the image and descriptor space respectively.

The approach outlined above leads to the following algorithm, illustrated in Figure 1:

Algorithm 1: Rank Minimization Based Inpaiting:

0. Data: A sequence of frames $\mathcal{I}_t$, $t = \{1, T\}$. The target is occluded/corrupted in $r \leq t \leq s$.
1. Extract the target z_t in the unoccluded frames, $1 \leq t < r$ and $s < t \leq T$. This step can be accomplished for example using the approach proposed in [23].
2. At each time t, $t = \{1, \cdots, r, s, \cdots, T\}$, map the pixel targets, z_t to a point y_t in a low dimensional manifold, using a Nonlinear Dimensionality Reduction (NDR) method such as Local Linear Embeddings (LLE) [35, 11, 21], described in the Appendix.
3. Find the descriptor values $\{y_r, \cdots, y_s\}$ corresponding to missing/occluded frames by minimizing the rank of the corresponding Hankel matrix.
4. Reconstruct the pixels z_t from the points y_t using a nonlinear mapping that approximately inverts the projection onto the low dimensional manifold. This inverse mapping can be obtained by using Radial Basis Functions (RBF) [25, 21], trained using data from the unoccluded frames.
5. Use a robust tracking method such as the approach proposed by Camps *et al.* [5] to estimate the position of the centroid of the target c_t in the occluded frames.
6. Inpaint the reconstructed appearance in these positions to accomplish the final result.

Example 1: Quasi Periodic Motion. Consider the sequence shown at the top of Figure 2, consisting of 78 frames captured with a stationary camera. The goal is to unocclude the person walking in the background, occluded in frames 37 through 49.

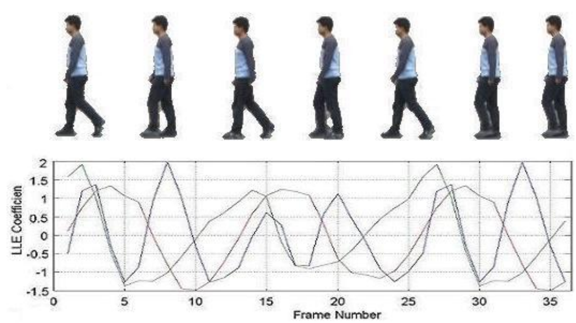


Figure 3. Mapping from image space to descriptor space via LLE. Top: sample appearances from walking motion. Bottom: Corresponding representation on a three dimensional manifold obtained via LLE.

The results of applying Algorithm 1 are shown in Figure 2, with the details of steps 2 (LLE learning) and 3 (interpolation) illustrated in Figures 3 and 4 respectively. As shown there, in this case the proposed algorithm resulted in virtually perfect inpaiting.

Example 2: Non Periodic Motion 1. Consider now the non–periodic motion shown in Figure 5. The total lenght of

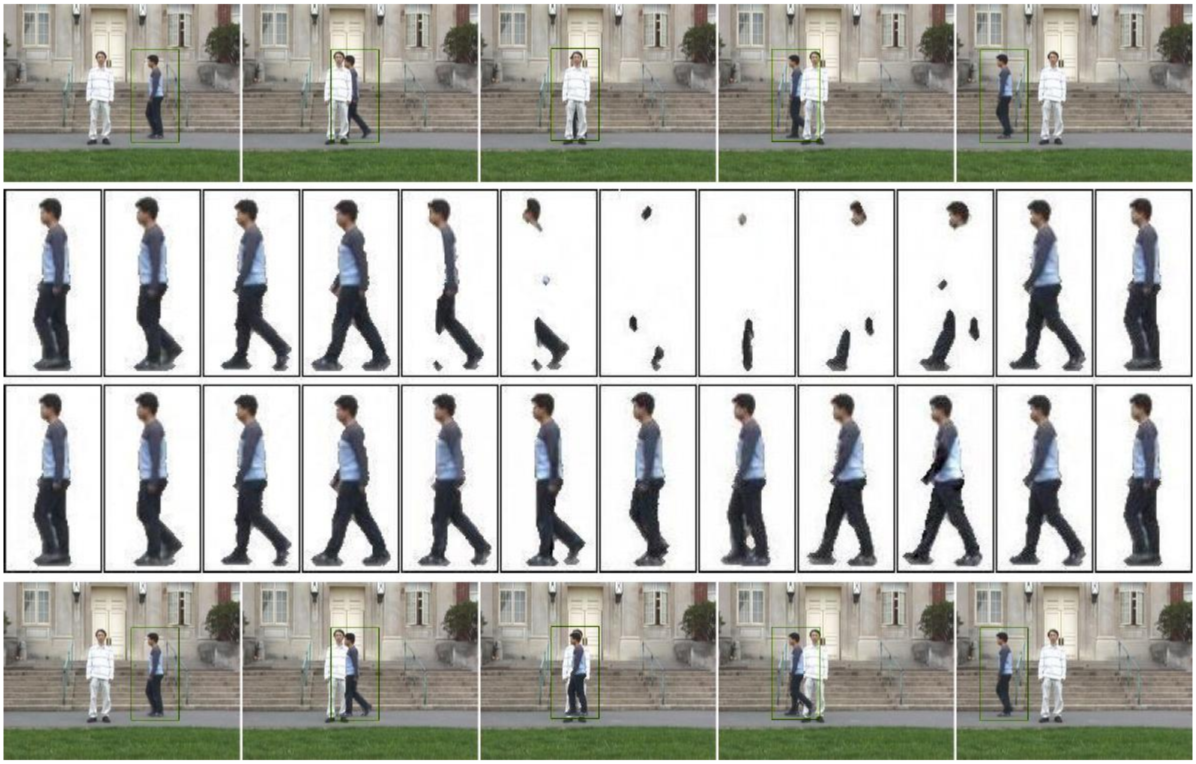


Figure 2. Applying Algorithm 1 to Inpaint Quasi Periodic Motion. Top row frames 17, 38, 43, 48, and 54 of the original sequence (from left to right). Second row: Learning the low dimensional manifold using unoccluded frames. Third row: Reconstructed target using Radial Basis Functions. Bottom row: Resulting inpainted sequence.

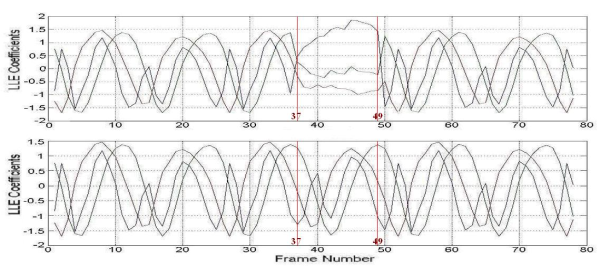


Figure 4. Top: low dimensional (3-D) representation of a walking sequence on an LLE manifold; frames 37-49 have occlusion. Bottom: Descriptor sequence interpolated via rank minimization.

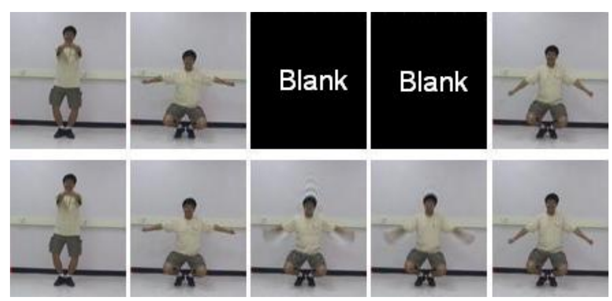


Figure 5. Video inpainting of target with non-periodic motion. From the top to bottom: original sequence and inpainting results

this sequence is 25 frames, with frames 16–18 missing. The results of applying Algorithm 1 are shown at the bottom of the figure. In this case, the algorithm achieved good reconstruction, with some blurring appearing around the hands, due to a relative short training sequence and reconstruction error from the Radial Basis Functions. It is worth emphasizing the fact that the missing information is not available anywhere else in the sequence, and thus cannot be reconstructed by finding suitable patches. Rather, reconstruction is made possible in our framework by simultaneous exploitation of spatial correlation and temporal dynamics.

Example 3: Non Periodic Motion 2.

Consider the sequence shown at the top of Figure 6, consisting of 243 frames of a sequence of a toy Robot on top of a slowing down turntable, with occlusion in frames 76–105. The inpainting results obtained using Algorithm 1 and graphcut [18] are shown in the middle and bottom of the figure, respectively. As illustrated there, although both approaches lead to smooth reconstruction, graphcut results in an incorrect reconstruction of the pose of the object. The original and interpolated LLE sequences are shown in Figure 7 (a) and (b).

Example 4: Moving Camera. This example illustrates the ability of Algorithm 1 to deal with sequences captured us-

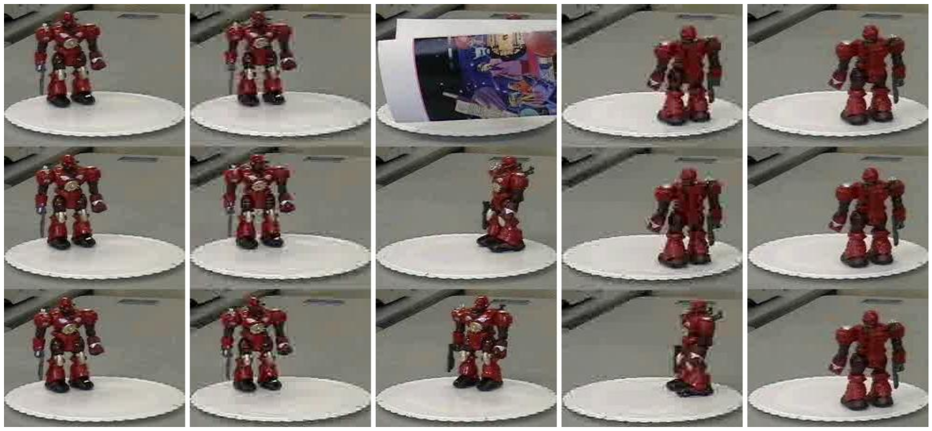
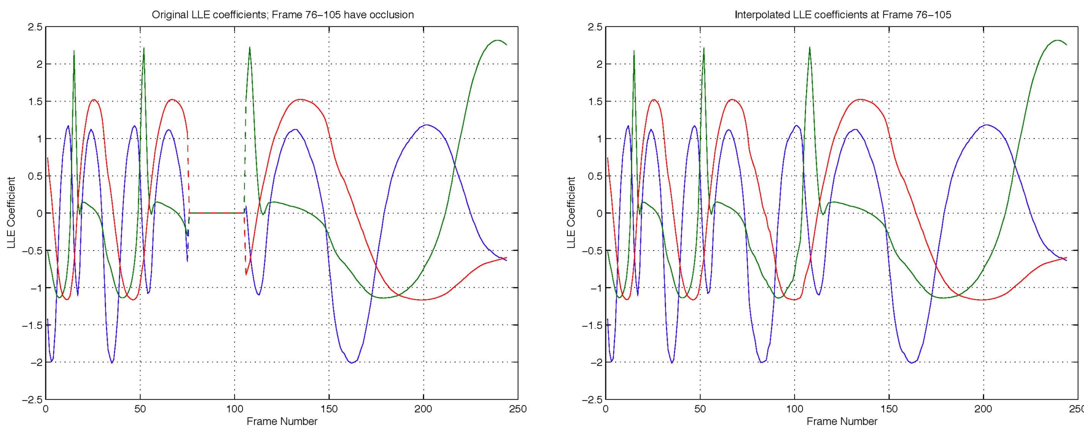


Figure 6. Applying Algorithm 1 and graphcut to inpaint non-periodic motion. Top row frames 75, 76, 90, 105, and 106 of the original sequence (from left to right). Second row: Inpainting result via Algorithm 1. Third row: Inpainting result via graphcut.



(a) (b)

Figure 7. (a) The original LLE sequence, where frame 76-105 have occlusion. (b) The interpolated LLE sequence obtained via rank minimization.

ing a moving camera, possibly undergoing both rotation and translation, as long as the dynamics of the camera motion can be described by a linear time invariant model (which can be learned from the data). In this case, inpainting is accomplished by separately learning the dynamics of the foreground and background descriptors. In the case of the foreground, the dynamics of the moving camera are absorbed into the dynamics of the descriptors. In the case of the (stationary) foreground[1], estimating descriptor evolution is effectively equivalent to estimating camera motion.

Figure 8 shows an application of the ideas outlined above. For the sake of simplicity, here we have chosen as descriptors the position of the 6 joints indicated in red, and used rank minimization to estimate their position (blue dots). Reconstruction was accomplished by finding the best match (in the minimal distance sense) to these descriptors in the unoccluded frames and cloning the corresponding target.

[1] a class of dynamic backgrounds will be addressed in the next section.

Example 5: Dynamic Texture. In this example we consider the problem of inpainting dynamic textures, e.g. sequences whose frames are relatively unstructured, but possessing some overall stationary properties. The proposed approach can be applied to these cases by first extrapolating the values of the missing Fourier descriptors[1] and then recreating the missing frames via a simple inverse FFT. In this case, the inpainting algorithm consists of the following steps: (i) Find the Fourier descriptors in the unoccluded frames by performing a 2-D DFT of each frame, (ii) Estimate the values of these descriptors in the missing frames by minimizing the rank of the corresponding Hankel matrix; and (iii) Recreate the missing frames by simply taking the inverse FFT of these descriptors.

The same ideas can also be used to restore frames having both, *a dynamically textured background*, and *moving objects in the foreground*. An example of this situation is shown in Fig. 9, where frames 10 and 24 are missing. Here the 1st row is the original sequence, and the 2nd row is the

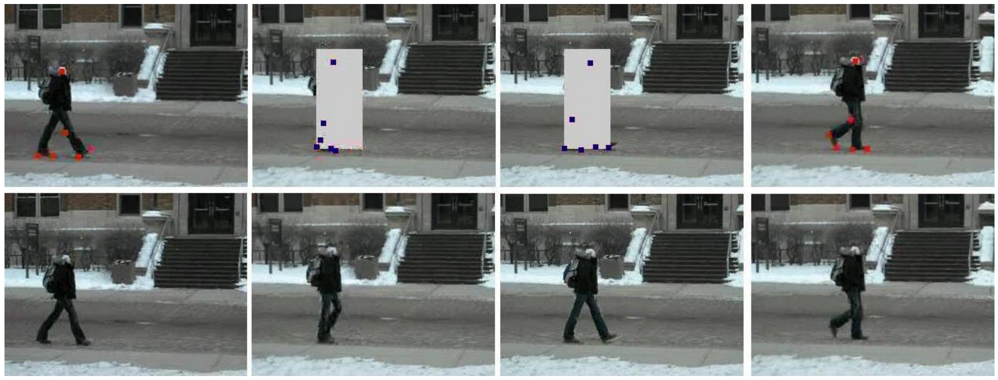


Figure 8. Inpainting results for the case of a moving camera. from left to right: frame 7, 11, 13, and 17, where frame 11-13 have occlusions. Top: descriptor definition (red points) and estimation (blue points) via rank minimization; Bottom: inpainting results.

Figure 9. Inpainting results of moving person in front of a dynamic texture background. From left to right: frame 11, 10, 17, 24, and 31, where frame 10 and 24 are missing. First row: original sequence. Second row: restoration of the moving person (foreground layer). Third row: restoration of the river dynamic texture (background layer). Fourth row: final inpainting results.

moving person sequence generated by the process outlined Example 1. Note that the Fourier descriptors can no longer be used for background restoration, since the black region will significantly affect the Fourier transform. Following an idea from [8], we considered the gray value along the frame sequence corresponding to each pixel as a descriptor sequence. Since the object is moving, the number of missing descriptors in each sequence is far smaller than the amount of known ones and thus the computational complexity of the overall process remains moderate. The reconstructed background is shown in the 3rd row of Fig. 9, and the final restoration result is shown in the 4th row.

4. Conclusions and Directions for Further Research

In this paper we propose a new approach for video inpainting based on the idea of completing the missing information in such a way that it leads to the simplest (e.g. lowest order) model explaining the spatio–temporal evolution of the pixels in the sequence of frames. This was accomplished by exploiting spatial correlations to map suitable regions of the image to descriptors in a lower dimensionality manifold, where missing values were interpolated by solving a rank minimization problem. Finally, the missing portions of the image are reconstructed from these interpolated descriptors by using a radial–basis functions based expansion, trained using data from the uncorrupted frames. When compared with existing approaches, the combination spatial dimensionality reduction/temporal dynamic modelling allows for

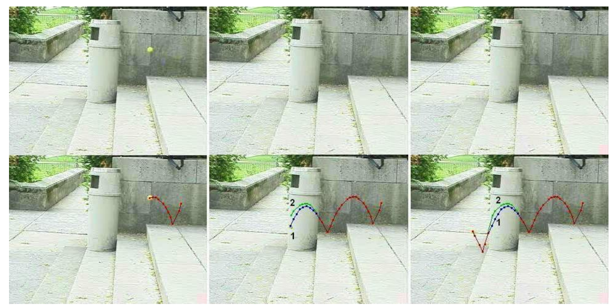


Figure 10. Top: Jumping ball sequence. Bottom: Trajectory reconstruction result. Blue line: Rank Minimization Based. Green line: PDE based approach [23]

1. Exploitation of the underlying temporal correlation to handle cases where the motion of the target is not quasi–periodic (as illustrated in Example 2, section 3). In addition this leads to smoother results in the case of severe occlusion. This effect is illustrated in Figure 10, comparing the results of rank minimization against the algorithm proposed in [23]. In addition to yielding better reconstruction of the missing part of the trajectory, rank minimization is computationally simpler, since it involves solving a Linear Matrix Inequality optimization problem rather than partial differential equations.

2. Smoothly interpolating missing frames in dynamic textures, with moderate computational complexity. For comparison the approach proposed in [8], has higher computational cost, since it uses dynamical information at the pixel level, and albeit it can model and extrapolate entire sequences, is not well suited for smoothly interpolating missing frames.

The main shortcoming of the proposed approach is that it cannot currently deal with cases involving scaling changes or deformations. Efforts are currently underway to address this problem by using results from nonlinear identification.

Appendix: Technical details related to Nonlinear Embeddings and Linear Systems theory

4.1. Locally Linear Embeddings

In this section we provide a brief description of a nonlinear dimensionality reduction method, Locally Linear Embeddings (LLE), that preserves local neighborhoods [35]. This method has been successfully used to model and learn human appearance changes in [11, 21]. Given T frames of a sequence, denote by z_t the vector obtained by stacking the pixels of the target at frame t. The goal is to associate to each vector z_t a point y_t in a lower dimensional manifold, e.g. $\dim(y_t) \ll \dim(z_t)$, while preserving the local structure. This is accomplished proceeding as follows:

1. Compute the neighbors set z_j of each data point z_t.
2. Calculate the weighs w_{ij} that best reconstruct z_t from its neighbors
   z_j, by minimizing $\sum_t \|z_t - \sum_j w_{tj} z_j\|^2$,
   where $\sum_j w_{ij} = 1$.
3. Compute the vector y_i best reconstructed by the weights w_{ij},
   minimizing $\sum_t \|y_t - \sum_j w_{tj} y_j\|^2$, subject to the constraint $\frac{1}{N}\sum_i y_i y_i^T = I$.

4.2. Model Order Estimation via Hankel Matrices:

In this section, we briefly discuss some results of Linear Systems Theory related to the proposed approach and give an intuitive explanation of the main idea behind the algorithm. We begin by recalling the following result, relating the rank of the Hankel matrix with the minimal order of the corresponding model [38] [31].

Fact 1: Denote by f_k the impulse response sequence of the model (1). Then, the minimum order n_o of the linear time invariant (LTI) system that has this impulse response is related to the rank of the corresponding Hankel matrix by:

$$\text{rank}[H_f] \leq n + n_0, \quad n \geq n_0. \tag{3}$$

Moreover, if the impulse input is sufficiently rich, e.g. it excites all the modes of the system, and $n >> n_0$, the equality holds. $\square$.

This result suggest the following procedure (loosely related to subspace Identification methods) to interpolate missing elements in the output sequence: Given a partial sequence $\mathbf{f}_g = \{f_1, \cdots, f_q, f_{q+r}, \cdots, f_n\}$, $f_i \in R$, of the impulse response of a LTI system LTI system with McMillan degree $n_o \ll n$, find the missing elements by solving the following optimization problem:

$$\vec{f}_o = \underset{\mathbf{f}_x}{argmin}\, \text{rank}[H_f]$$

where $\mathbf{f}_x = \{f_{q+1}, \ldots, f_{q+r-1}\}$ and H_f denotes the Hankel matrix associated with the entire sequence $\{f_i\}$.

Intuitively, the procedure above can be understood as simply indicating that the missing elements that best fit a given sequence are those that require adding the least number of states to the existing model in order to explain the new data. For instance, in the case of a sinusoidal wave, this approach will reconstruct the missing portion by first estimating its frequency and then adding points compatible with it, rather than adding to the model components with additional frequency content.

A potential difficulty with the approach outlined above is that rank minimization is generically NP–hard [36]. However, there currently exist efficient tractable relaxations to approach this problem. The one used in this paper, originally proposed in [13], recasts the rank minimization problem into the following (convex) trace minimization form:

$$\begin{array}{ll} \text{minimize} & \text{rank}[X] \\ \text{subject to} & X \in \mathcal{C} \end{array} \implies$$

$$\text{TMP}: \quad \begin{array}{ll} \text{minimize} & Tr(Y) + Tr(Z) \\ \text{subject to} & \begin{bmatrix} Y & X \\ X^T & Z \end{bmatrix} \geq 0 \\ & X \in \mathcal{C} \end{array}$$

where $\mathcal{C}$ is a convex set, $X \in R^{m \times n}$, $Y = Y^T \in R^{m \times m}$ and $Z = Z^T \in R^{n \times n}$.

References

[1] B. Abraham, O. I. Camps, and M. Sznaier. Dynamic texture with fourier descriptors. In *Proceedings of the 4th International Workshop on Texture Analysis and Synthesis*, volume 1, pages 53–58. IEEE, 2005.

[2] M. Bertalmio, G. Sapiroand, V. Caselles, and C. Ballester. Image inpainting. In *Proceedings of SIGGRAPH 2000*, volume 1, pages 417–424. ACM, 2000.

[3] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher. Simultaneous structure and texture image inpainting. *IEEE Transactions on Image Processing*, 12(8):882–889, 2002.

[4] W. Cai, J. Wu, and A. Chung. Shape-based image segmentation using normalized cuts. In *Proceedings of ICIP 2006*, volume 1, pages 1101–1104. IEEE, 2006.

[5] O. Camps, H. Li, M. C. Mazzaro, and M. Sznaier. A caratheodory-fejer approach to robust multiframe tracking", computer vision. In *Proceedings of ICCV 2003*, volume 2, pages 1048–1055. IEEE, 2003.

[6] V. Cheung, B. J. Frey, and N. Jojic. Video epitomes. In *Proceedings of CVPR 2005*, volume 2, pages 42–49. IEEE, 2005.

[7] A. Criminisi, P. Perez, and K. Toyama. Region filling and object removal by exemplar-based image inpainting. *IEEE Transactions on Image Processing*, 13(9):1200–1212, 2004.

[8] G. Doretto, A. Chiuso, Y. N. Wu, and S. Soatto. Dynamics textures. *International Journal of Computer Vision*, 51(2):91–109, 2003.

[9] I. Drori, D. Cohen-Or, and H. Yeshurun. Fragment-based image completion. *ACM Transactions on Graphics (TOG)*, 22(3):303–312, 2003.

[10] A. Efros and T. Leung. Texture synthesis by non-parametric sampling. In *Proceedings of ICCV 1999*, volume 2, pages 1033–1038. IEEE, 1999.

[11] A. Elgammal and C. S. Lee. Inferring 3d body pose from silhouettes using activity manifold learning. In *Proceedings of CVPR*, volume 2, pages 681–688. IEEE, 2004.

[12] M. J. Fadili and J. L. Starck. Em algorithm for sparse representation-based image inpainting. In *Proceedings of ICIP 2005*, volume 2, pages 61–64. IEEE, 2005.

[13] M. Fazel, H. Hindi, and S. Boyd. Rank minimization and applications in system theory. In *Proceedings of American Control Conf. 2004*, volume 4, pages 3273–3278. AACC, 2004.

[14] J. Jia, W. Tai-Pang, T. Yu-Wing, and T. Chi-Keung. Video repairing: inference of foreground and background under severe occlusion. In *Proceedings of CVPR 2004*, volume 1, pages 364–371. IEEE, 2004.

[15] Y. T. Jia, S. M. Hu, and R. R. Martin. Video completion using tracking and fragment merging. In *Visual Computer*, volume 21, pages 601–610. Springer-Verlag, 2005.

[16] A. Kokaram. Practical mcmc for missing data treatment in degraded video. In *Proceedings of ECCV workshop on Statistical methods in video processing 2002*, volume 2352, pages 85–90. European Vision Society (EVS), 2002.

[17] A. C. Kokaram, R. D. Morris, W. J. Fitzgerald, and P. Rayner. Interpolation of missing data in image sequences. *IEEE Trans. on Image Processing*, 4(11):1509–1519, 1995.

[18] V. Kwatra, A. Schodl, I. Essa, and A. Bobick. Graphcut textures: image and video synthesis using graph cuts. In *SIGGRAPH*, 2003.

[19] V. Kwatra, A. Schodl, I. Essa, G. Turk, and A. Bobick. Semidefinite programming for discrete optimization and matrix completion problems. *ACM Transactions on Graphics*, 22(3):277–286, 2003.

[20] A. levin, A. Zomet, and Y. Weiss. Learning how to inpaint from global image statistics. In *Proceedings of ICCV 2003*, volume 1, pages 301–312. IEEE, 2003.

[21] H. Lim, V. I. Morariu, O. I. Camps, and M. Sznaier. Dynamic appearance modeling for human tracking. In *Proceedings of CVPR*, volume 1, pages 751–757. IEEE, 2006.

[22] S. Masnou and J. Morel. Level-lines based disocclusion. In *Proceedings of ICIP 1998*, volume 3, pages 259–263. IEEE, 1998.

[23] K. A. Patwardhan, G. Sapiro, and M. Bertalmio. Video inpainting of occluding and occluded objects. In *Proceedings of ICIP 2005*, volume 2, pages 69–72. IEEE, 2005.

[24] R. S. Pena and M. Sznaier. *Robust Systems Theory and Applications*. Wiley & Sons, Inc., 1998.

[25] T. Poggio and F. Girosi. Networks for approximation and learning. In *Proceedings of IEEE*, volume 78, pages 1481–1497. IEEE, 1990.

[26] S. Roth and M. J. Black. Fields of experts: a framework for learning image priors roth. In *Proceedings of CVPR 2005*, volume 2, pages 860–867. IEEE, 2005.

[27] H. Shen and J. Coughlan. Finding text in natural scenes by figure-ground segmentation. In *Proceedings of ICPR 2004*, volume 4, pages 113–118. IEEE, 2004.

[28] T.-K. Shih, N.-C. Tang, and W. Lee. Video inpainting and implant via diversified temporal continuations. In *Proceedings of ACM MULTIMEDIA 2006*, volume 1, pages 133–136. ACM, 2006.

[29] T. Shiratori, Y. Matsushita, X. Tang, and S. Kang. Video completion by motion field transfer. In *Proceedings of CVPR 2006*, volume 1, pages 411–418. IEEE, 2006.

[30] E. Simoncelli and J. Portilla. Texture characterization via joint statistics of wavelet coefficient magnitudes. In *Proceedings of ICIP 1998*, volume 1, pages 62–66. IEEE, 1998.

[31] N. K. Sinha and B. Juszta. *Modeling and Identificatio of Dynamic Systems*. Van Nostrand Reinhold Co., 1983.

[32] J. L. Starck, M. Elad, and D. L. Donoho. Image decomposition via the combination of sparse representations and a variational approach. *IEEE Transactions on Image Processing*, 14(10):1570–1582, 2005.

[33] J. Sun, L. Yuan, J. Jia, and H. Y. Shum. Image completion with structure propagation. *ACM Transactions on Graphics*, 12(3):861–868, 2005.

[34] M. Szummer and R. W. Picard. Temporal texture modeling. In *Proceedings of Internationl Conference on Image Processing*, volume 3, pages 823–826. IEEE, 1996.

[35] S. T. Toweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[36] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, 1996.

[37] Y. Wexler, E. Shechtman, and M. Irani. Space-time video completion. In *Proceedings of CVPR 2004*, volume 1, pages 120–127. IEEE, 2004.

[38] C. M. Woodside. Estimation of the order of linear systems. *Automatica*, 7(6):727–733, 1971.

[39] Y. Zhang, J. Xiao, and M. Shah. Motion layer based object removal in videos. In *Proceedings of WACV 2005*, volume 1, pages 516–521. IEEE, 2005.