Video Denoising based on Sparse Transformation and Low Rank Matrix Completion

Zhijie Lin [1,2, a], Xiaohua Li[3,b] , Zhijun Sun[4,c] and Ping Zeng[5,d]

[1] *Zhejiang University of Science and Technology*
[2]*The Department of Computer Science and Technology, Zhejiang University*
[3]*The Information Center of Zhejiang University*
[4]*The Center of Digital Cultural Heritage of Dunhuang Research Institute*
[5]*Science and Technology Research Institute of China Petroleum Group*
[a]*bytelin@qq.com,* [b]*xiaohua@zju.edu.cn,* [c]*sunfoto@qq.com,* [d]*zping@petro china.com.cn*

Abstract

In this paper, we combine two powerful tools to handle the video denoising problem: one is an effective video denoising method based on highly sparse signal representation in local 3D transform domain, and the other is a low-rank matrix completion based video denoising method. Similarly, in our algorithm, a noisy video is processed in block-wise manner and for each processed block we form a 3D data array that we call "group" by stacking together blocks found similar to the currently processed one. "Collaborative filtering" exploits the correlation between grouped blocks and the corresponding highly sparse representation of the true signal in the transform domain. By employ low-rank matrix completion method in our framework, our technique is also robust to different types of noise, such as Gaussian additive noise and impulsive noise. Experiments demonstrate that our techniques produce state-of-the-art results for video denoising applications.

Keywords: video denoising, low-rank matrix completion, transform domain, block matching.

Introduction

With today's advances in sensor design, the image/video is relatively clean for high-end digital cameras at low sensitivities, but it remains noisy for low cost cameras at high sensitivities, e.g., low light condition, high ISO setting and high speed rate. The problem of removing image noise is still of acute and in fact growing importance with the prevalence of webcams and mobile phone cameras. In general, video data tend to be more noisy than single image due to high speed capturing rate of video camera. Video denoising aims at efficiently removing noise from all frames of a video by utilizing information in both spatial and temporal domains. Such an integrated approach is more optimal than independently applying a single-image denoising method on each frame of the video, as there exist high temporal redundancies in a video compared to a single image.[1] These methods typically utilize both the sparsity and the statistical properties of a multiresolution representation as well as the inherent correlations between frames in temporal dimension. A recent denoising strategy, the non-local spatial estimation [2], has also been adapted to video denoising [3]. In this approach, similarity between 2D patches is used to determine the weights in a weighted averaging between the central pixels of these patches. For image denoising, the similarity is measured for all patches in a 2D local neighborhood centered at the currently processed coordinate. For video denoising, a 3D such neighborhood is used. The effectiveness of this method depends on the presence of many similar true-signal blocks.[4]

Based on the same assumption as the one used in the non-local estimation, i.e. that there exist mutually similar blocks in natural images, in [5] the authors proposed an image denoising method. There, for each processed block, we perform two special procedures: grouping and collaborative filtering. Grouping finds mutually similar 2D blocks and then stacks them together in a 3D array that we call group. The benefit of grouping highly similar signal fragments together is the increased correlation of the true signal in the formed 3D array. Collaborative filtering takes advantage of this increased correlation to effectively suppress the noise and produces estimates of each of the grouped blocks. They showed [5] that this approach is very effective for image denoising.

In this paper, we apply the concepts of grouping and collaborative filtering to video denoising. Grouping is performed by a specially developed predictive-search block matching technique that significantly reduces the computational cost of the search for similar blocks. We employ a two-step video-denoising algorithm proposed in [4] where the predictive search block-matching is combined with collaborative hard thresholding in the first step and with collaborative Wiener filtering in the second step. In order to enhance the robustness of the algorithm in processing denoising problems with multiple sources of noises, our algorithm is derived with minimal assumptions on the statistical properties of image noise. The basic idea is to convert the problem of removing noise from the stack of matched patches to a low rank matrix completion problem, which can be efficiently solved by minimizing the nuclear norm of the matrix with linear constraints.[1]

Related Work

There have been an abundant research literature on image denoising methods. In this section, we will only discuss the most related denoising techniques. There have been abundant research literature on image denoising methods. In this section, we will only discuss the most related denoising techniques [4,6,7]. Although differing from details, these method are built on the same methodology which essentially groups the similar patches together followed by a collaboratively filtering. Take the well-known BM3D [5] as a sample. In BM3D, similar image blocks is stacked in a 3D array based on the L2 norm distance function between different patches. Then a shrinkage in 3D transform domain such as wavelet shrinkage or Wiener filter are applied on the 3D block stack. The denoised image is then synthesized from denoised patches after inversing 3D transform. The result can be further improved by iteratively doing grouping and collaboratively filtering.

Video denoising is different from single image denoising as video sequences usually have very high temporal redundancy which should be effectively used for better performance (e.g., [8-10]). The basic idea of patch-based image denoising

can also be applied on the video by matching similar patches both within the image and over multiple images. The concept of BM3D is generalized to video denoising in [4] by using a predictive search block-matching over time and combined with collaborative Wiener filtering on patch stacks. In [11], a more robust patch matching are proposed by using the depth as a constraint in the matching process and the patch stack is denoising by both PCA (principle component analysis) and Tensor analysis. The idea of sparse coding in a patch dictionary has also been applied on video denoising, where the denoised image patches are found by seeking for the sparsest solution in a patch dictionary. Among these patch-based video denoising techniques, most assume data noise is only additive i.i.d. Gaussian noise (e.g., [4, 12, 13]). The image noise mixed with both Gaussian noise and Poisson shot noise are considered in [14]. Regarding impulsive noise, there have been many research works on removing impulsive noise (or salt-and pepper noise) from a single image.

Our Approach

Block matching and grouping

A straightforward approach is to use a fixed-size 3D search neighborhood for the grouping by block-matching. However, capturing blocks of a moving object across many frames requires large spatial dimensions of such search neighborhood. On the one hand, using large sizes imposes a rather high complexity burden, and on the other hand, using small ones results in unsatisfactory grouping and poor denoising results.

In this part we employ the strategy utilized in [4] to solve our problem. In order to efficiently capture blocks that are part of objects which move across subsequent frames, we propose to use predictive-search block-matching, an inductive procedure that finds similar (matching) blocks by searching in a data adaptive spatio-temporal subdomain of the video sequence. For a given reference block located at $X = (x_1, x_2, t_0)$, when using a temporal window of $2*N_{FR} + 1$ frames, the predictive search block-matching comprises the following steps.

Step 1: Starting with frame t_0, an exhaustive-search block matching is performed in a non-adaptive $N_S * N_S$ neighborhood centered about (x_1, x_2). The results are the spatial locations of the N_S blocks (within this neighborhood) which exhibit highest similarity to the reference one. These locations are collected in the set $S^{t_0} \in Z^3$.

Step 2: The predictive search in frame t_0+k, $0<|k|<N_s$, is defined inductively based on the matching results from the previously processed frame $t_0 + k\text{-sign}(k)$, i.e. from the preceding frame for $k > 0$ or from the subsequent frame for $k<0$. This search for similar blocks takes place within the union of $N_{FR} * N_{FR}$ neighborhoods centered at the spatial coordinates of the previously found locations $x \in S^{t_0+k-\text{sign}(k)}$. That is, these locations predict where similar blocks are likely to be present in the current frame (i.e. frame t_0+k) and thus one can afford to have $N_{FR} < N_s$. The result for the current frame are the N_B locations of the blocks that exhibit highest similarity to the reference one; they are collected in the set S^{t_0+k}.

After performing the predictive-search block-matching for all of the frames t_0+k for t_0 $k=-N_{FR}, \ldots, N_{FR}$, we form a single set $S_x \in Z^3$ that contains at most N_2 of all

$$X' \in \bigcup_{K=-N_{FR}}^{N_{FR}} S^{t_0+k}$$

that have the smallest corresponding block-distances to the reference block, which distances should also be smaller than a predefined threshold λ. A group is later formed by stacking together blocks located at $x \in S_x$. The exact ordering of the blocks within the 3D groups is not important, as shown in [5]. In the worst case, no matching blocks are found and then the group will contain only one block since its distance to itself is zero and therefore x will always be included in S_x. [4]

Except for the frame t_0 in the procedure presented above, the spatial search neighborhoods are data-adaptive as they depend on previously matched locations.

This adaptivity can be interpreted as following the motion of objects across frames.

Denoising patch matrix

By applying the two-stage algorithm described above on each patch of inputed image frames, we can effectively remove most noises from all patches[1]. The last step is to synthesize the denoised image from these denoised patches. In our implementation, the image patches are sampled with overlapping regions. Thus, each pixel is covered by several denoised patches. Then, the value of each pixel in images is determined by taking the average of denoised patches at this pixel which will suppress the possible artifacts in the neighborhood of the boundaries of patches.[1]

Conculusion

In this paper, we combine two powerful tools to handle the video denoising problem: one is an effective video denoising method based on highly sparse signal representation in local 3D transform domain, and the other is a low-rank matrix completion based video denoising method. This work is supported in part by Zhejiang Provincial projects (2014C33069), and the National Key Technology R&D Program projects (2013BAH27F01, 2013BAH27F02, 2012BAH43F06).

References

[1] Ji, H., Liu, C., Shen, Z., Xu, Y, "Robust video denoising using low rank matrix completion," CVPR 2010.

[2] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," Multisc. Model. Simulat., vol. 4, no. 2, pp. 490-530, 2005.

[3] A. Buades, B. Coll, and J. Morel, "Denoising image sequences does not require motion estimation," in Proc. IEEE Conf. Adv. Video Signal Based Surveil., AVSS, Palma de Mallorca, Spain, September 2005, pp. 70-74.

[4] Kostadin, D., Alessandro F., and Karen E., "Video denoising by sparse 3D transform-domain collaborative filtering," EUSIPCO 2007.

[5] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering", IEEE Trans. Image Process., vol. 16, No. 8, 2007.

[6] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," IEEE Trans. Image Processing, 15(12), 2006.

[7] A. Foi, V. Katkovnik, and K. Egiazarian, "Pointwise shape adaptive DCT for high-quality denoising and deblocking of grayscale and color images," IEEE Trans. Image Processing, 16(5), 2007.

[8] J. Chen and C. K. Tang, "Spatio-temporal markov random field for video denoising," CVPR 2007.

[9] H. Hwang and R. A. Haddad, "Adaptive median filters: New algorithms and results," IEEE Trans. on Image Processing, 4:499–502, 1995.

[10] V. Vaish, M. Levoy, R. Szeliski, C. L. Zitnick, and S. B. Kang, "Reconstructing occluded surfaces using synthetic apertures: Stereo, focus and robust measures," CVPR 2006.

[11] L. Zhang, S. Vaddadi, H. Jin, and S. K. Nayar, "Multiple view image denoising," ICCV 2009.

[12] D. Rusanovskyy, K. Dabov, and K. Egiazarian, "Moving-window varying size 3D transform-based video denoising," in Proc. Int. Workshop on Video Process. and Quality Metrics, Scottsdale, Arizona, USA, January 2006.

[13] C.-C. Wan and C.-H. Hsieh, "An efficient fractal image coding method using interblock correlation search," IEEE Trans. Circuits Syst. Video Tech., vol. 11, no. 2, pp. 257-261, February 2001.

[14] E. Balster, Y. Zheng, and R. Ewing, "Combined spatial and temporal domain wavelet shrinkage algorithm for video denoising," IEEE Trans. Circuits Syst. Video Tech., vol. 16, no. 2, pp. 220ñ230, February 2006.

[15] Lin, Z., Chen, M., Ma, Y, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," UIUC Technical Report UILUENG-09-2215, 2010.

[16] M. Protter and M. Elad. "Image sequence denoising via sparse and redundant representations," In IEEE Trans. Image Processing, 2009.