Video Texture Smoothing Based on Relative Total Variation and Optical Flow Matching

Huisi Wu, Songtao Tu, Lei Wang and Zhenkun Wen

Abstract Images and videos usually express perceptual information with meaningful structures. By removing fine details and reserving important structures, video texture smoothing can better display the useful structural information, and thus is significant in the video understandings for both human and computers. Compared with the image texture smoothing, video texture smoothing is much more challenging. This paper proposes a novel video texture smoothing method through combining existing Relative Total Variation (RTV) and optical flow matching. By considering both special relationship and color/gradient similarity between adjacent frames, we build an optimization framework with two novel regularization terms and solve the smoothed video texture via iterations. Convincing experiment results demonstrate the effectiveness of our method.

Keywords Texture · Structure · Smoothness · Relative total variation · Optical flow · Probability model

1 Introduction

Video contents usually contain more or less natural or artificial textures, such as the grasses on the grounds or repeated patterns on the buildings. Such textural contents make the video more attractive. However, humans normally perceive video content by capturing the meaningful large scale structures. Thus, video texture smoothing becomes extremely important in computer vision and video pattern recognition, especially for video abstraction and object tracking.

H. Wu · S. Tu · L. Wang · Z. Wen (✉)
College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China
e-mail: wenzk@szu.edu.cn

Z. Wen and T. Li (eds.), *Practical Applications of Intelligent Systems*, Advances in Intelligent Systems and Computing 279,
DOI: 10.1007/978-3-642-54927-4_2,

To smooth the fine texture from image, most existing methods define the texture region based on total variation of the gradient fields [1, 2], and use an optimization algorithm to filter out fine details. Recently, Xu et al. [3] proposed a method for texture smoothing based on Relative Total Variation (RTV), where better smoothed effects can be obtained. Compared with the studies carried out on the image texture smoothing, video smoothing is more challenging and none of the existing methods can obtain favorable results. For video smoothing, not only the main structures should be persevered in each frame image, but also the correspondences between frames should be considered to solve stable-smoothed video results.

In this paper, we propose a novel video smoothing method based on RTV and optical flow matching. Since RTV is a well-validated method in the separation of texture and structure, it provides better effects compared with other methods. As an important method in the analysis of visual motion, optical flow method provides dense correspondences for large scale structures between two frames. By estimating the displacement for each pixel via optical flow, we formulate an inter-frame constraint to optimize video smoothing. Compared to existing methods, which only consider the video smoothing frame by frame, our method achieves a more stable smoothing effect by preserving the video continuity. Specifically, we applied optical flow to describe a constraint between two adjacent frames in both original and smoothed video. Based the combination of RTV and optical flow, we propose a graph model to express the spatial and temporal relation between the adjacent frames. The probability model is analyzed to form an object function. Finally, we obtain the smoothed video by solving the object function. Experimental results show that our method outperforms the existing RTV in terms of main structures preserving and texture details smoothing.

2 Related Work

Since texture smoothing can extract salient structures from textures, it has been a hot research topic in computer vision and pattern recognition. Rudin et al. [4] first proposed TV-L2 model, which can be easily expressed through fidelity data term and regularization term. Aujol et al. [5] had studied four kinds of TV models and come to the conclusion that TV-L2 had best effects in dealing images without knowledge of its texture model in advance, but it had some demerits in distinguishing the strong structural textures and edges. Farbman et al. [6] and Xu et al. [7], respectively, proposed WLS and L0 gradient minimization. Different from the regularization term and optimized procedure in TV-L2 model, their models have some disadvantages in dealing texture of different scales, and they still rely on gradient. Kass and Solomon [8] pointed out that local histogram-based filtering can well resist image details and maintain the structure edges at the time. But this method is not designed to process texture, and the direct usage of this method will not get desirable effects. Xu et al. [3] proposed the Relative Total Variation (RTV)

method, which is based on the TV-L2 model, and this method can better remove image texture, while maintaining the structural edge. However, we cannot directly use RTV to smooth video textures because RTV can only smooth singe image and does not consider the inter-frame relationship if it is used to smooth video.

Optical flow is state-of-the-art technique in calculating the pixel-wise correspondences between two images, especially for the neighboring frames in a video. The computation of optical flow field was first discussed by Horn and Schunck [9] in 1981. Based on the optical flow constraint equation, they supposed that the speed filed is smooth, and then they get a dense optical flow field. Nagel [10] comes up with the idea that second derivative can be used to deal with the optical flow field. With the application of an oriented smooth constraint to deal with the occlusion problem, the computation of the reciprocal value of second derivative can be easily impacted by the noise. Ghosal and Vanek [11] proposed that smooth constraints of different natures in speed fields can be used to compute the optical flow field. Lucase and Kanade [12] researched and developed many new concepts for dealing with shortcomings of previous models.

In this paper, we present a simple and yet effective model based on Relative Total Variation and optical flow matching to realize video texture smoothing. By adding spatial constraint to RTV based on optical flow, we can obtain a better result than only smoothing video frame by frame using RTV. The combination between RTV and optical flow turns out a nonlinear optimization problem. So we also propose an optimization solution to obtain the video smoothing results.

3 Approach

3.1 Model

For the video to be soothed, obviously, there are spatial and temporal relations between the original image and the smoothed one, which is shown in Fig. 1.

Suppose p_t is the pixel in the original image while s_t in the smoothed one, respectively. Both of them are in the same frame. The same position pixels in adjacent frame connect with each other through optical flow, for example, the connection between p_{t-1} and p_t as well as s_{t-1} and s_t. The original image and the smoothed one in the same frame connect with each other through Relative Total Variation [6], for example, the connection between s_t and p_t. From Fig. 1, it can be assumed that:

$$P(p_t|p_{t-1}) = \exp\left(-\|p_t - p_{t-1}\|^2\right) \tag{1}$$

$$P(s_t|s_{t-1}) = \exp\left(-\|s_t - s_{t-1}\|^2\right) \tag{2}$$

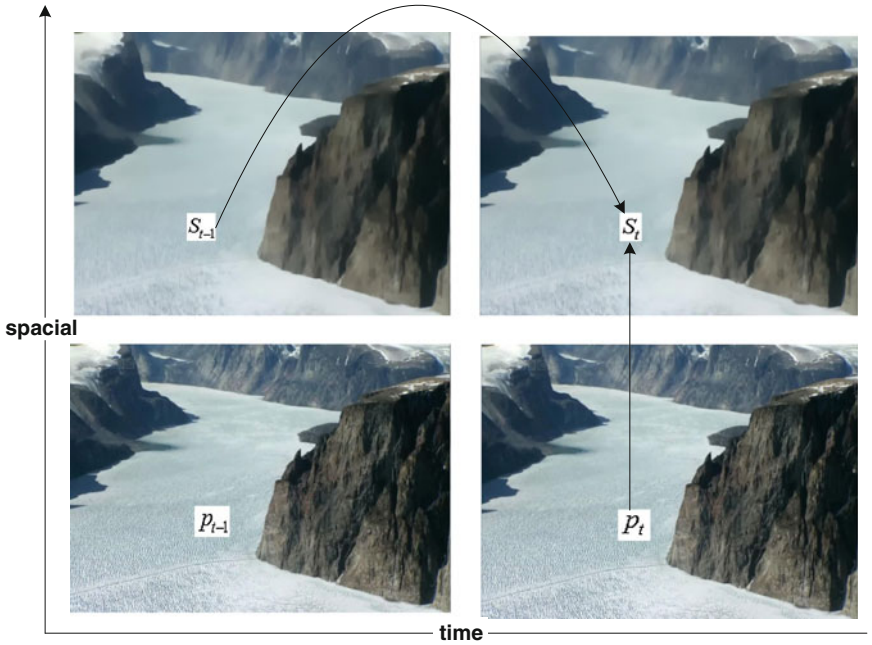

Fig. 1 From *left* to *right* display constrain relation between previous frame and the current frame, from *bottom* to *top* display constrain relation between original images and smoothed

$$P(p_t|s_t) = \exp(-u(s_t, p_t)) \tag{3}$$

Assume:

$$A = \|p_t - p_{t-1}\|^2; B = \|s_t - s_{t-1}\|^2; C = u(s_t, p_t)$$

It can be concluded:

$$\begin{aligned} P(p_t, s_t, p_{t-1}, s_{t-1}) &= P(p_{t-1})P(s_{t-1})P(s_t|s_{t-1})P(p_t|p_{t-1}, s_t) \\ &= P(p_t|p_{t-1})P(s_t|s_{t-1})P(p_{t-1})P(s_{t-1})P(p_t|s_t) \end{aligned} \tag{4}$$

From Bayes function:

$$P(p_t|s_t, p_{t-1}, s_{t-1}) = P(p_t, s_t, p_{t-1}, s_{t-1})/P(p_{t-1}, s_t, s_{t-1}) \tag{5}$$

From (4) and (5):

$$P(p_t|s_t, p_{t-1}, s_{t-1}) \propto P(p_t|p_{t-1})P(s_t|s_{t-1})P(p_t|s_t) \tag{6}$$

$$\log P(p_t|s_t, p_{t-1}, s_{t-1}) = -(A + B + C) \tag{7}$$

To maximize the probability, we should minimize the value of $A + B + C$. And we can obtain $(\Delta x, \Delta y)$ according the constraint conditions of optical flow

$$p_t(x, y) = p_{t-1}(x - \Delta x, y - \Delta y) \tag{8}$$

If t denotes any time, p_t (x,y) can be effectively estimated by $p_{t-1}(x - \Delta x, y - \Delta y)$. Thus we assumes:

$$p_t(x, y) - p_{t-1}(x - \Delta x, y - \Delta y) \propto N\left(0, \sigma^2\right) \tag{9}$$

Similarly, $s_t(x, y) - s_{t-1}(x - \Delta x, y - \Delta y)$ is also Gaussian distribution. As the original image and the smoothed one take Relative Total Variation as constraint condition [6], it can be said:

$$D = \sum_p (P_P - S_P)^2 + \lambda \cdot \left(\frac{D_x(p)}{\Phi_x(p) + \varepsilon} + \frac{D_y(p)}{\Phi_y(p) + \varepsilon} \right) \tag{10}$$

From (9) and (10), the objective function is finally expressed as:

$$\begin{aligned} &\arg\min_s \sum_p \left(p_t\left(x_p, y_p\right) - p_{t-1}\left(x_p - \Delta x, y_p - \Delta y\right)\right)^2 + \left(P_p - s_p\right)^2 \\ &+ \left(s_t\left(x_p, y_p\right) - s_{t-1}\left(x_p - \Delta x, y_p - \Delta y\right)\right) + \lambda \cdot \left(\frac{D_x(p)}{\Phi_x(p) + \varepsilon} + \frac{D_y(p)}{\Phi_y(p) + \varepsilon} \right) \end{aligned} \tag{11}$$

3.2 Numerical Solution

It is not easy to solve the target equation (11) as it is nonlinear. It contains three constraints: Item A, Item B, and Item C. Its solution thus cannot be obtained trivially. Therefore, we proposed a segmenting method to optimize the equation. Firstly, we use Item A and Item B to optimize it. Secondly, Item C will be applied to deal with the result got from last step. Finally, the whole target equation will get optimized. The Fig. 2 showed the flow chart. This method is clear in structure and simple in processing. Besides, after the process of optical flow matching, it will be very close to the result we want. Hence, the number of iterations will be reduced when we utilize Item C.

The steps involved are as follows:

(a) In Item A, $p_t(x, y)$ and $p_{t-1}(x, y)$ refer to the original image pixels, t means time. Δx and Δy can be accessed through $p_t(x, y)$ and $p_{t-1}(x, y)$ by optical flow matching. Item A can be viewed as a constant.

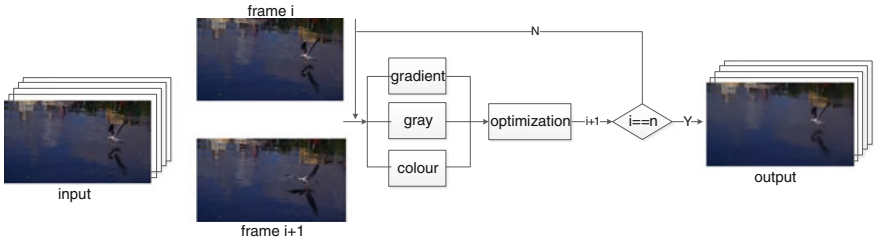

Fig. 2 The framework of our system

(b) Take Item C as the initial condition, through optimization, we get S_{t-1}.

(c) Optimize Item B through optical flow matching. The gray level of certain image point $w = (x, y)^T$ on time t is $I(x, y, t)$. After a timeslot of Δt, the correspond value for gray level is : $I(x + \Delta x, y + \Delta y, t + \Delta t)$. *Because the video image is continuous, the image changes slowly with x, y, t. With the Taylor series expansion and ignorance of the second-order infinitesimal items, we get*

$$\frac{\partial I}{\partial x}\mathrm{d}x + \frac{\partial I}{\partial y}\mathrm{d}y + \frac{\partial I}{\partial t} = 0 \tag{12}$$

It can be supposed that $u = \frac{\mathrm{d}x}{\mathrm{d}t}, v = \frac{\mathrm{d}y}{\mathrm{d}t}, I_x = \frac{\partial I}{\partial x}, I_y = \frac{\partial I}{\partial y}, I_t = \frac{\partial I}{\partial t}$ then function (12) can be rewritten as $I_x u + I_y v + I_t = 0$, $v_w = (u, v)$ is the optical flow for point w. Besides the gray value constancy assumption, we also consider color and gradient constancy assumption. After the above optimization, we get $S_t(x, y)$.

(d) Utilize Item C to constrain the equation to get the optimized solution. Taking the result got from step c as input, we apply the numerical solution in [3] to optimize the equation and get the final result.

4 Results

We collected a video datasets to evaluate our method, which contains 20 videos with different kinds of textures. Typical video smoothing results are as shown in Figs. 3 and 4. From the results shown in Fig. 3, we can clearly see that our method not only preserves the salient structures for each frame in the video, but also maintains the video consistency when smoothing a sequence of images, including color consistency and lighting consistency for the corresponding pixels along the time dimension.

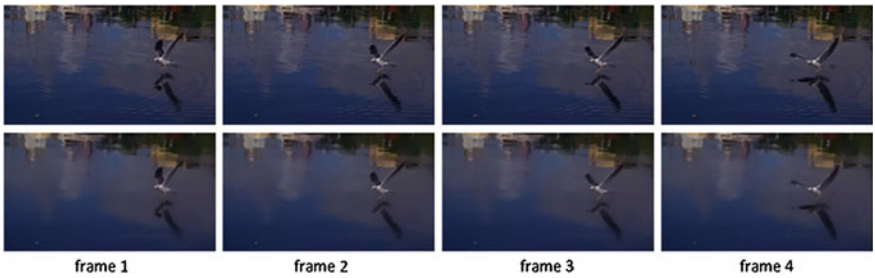

Fig. 3 Video smoothing results using our method. The *top row* is the original video from frame 1–4, and the smoothed results are shown in the *bottom row*

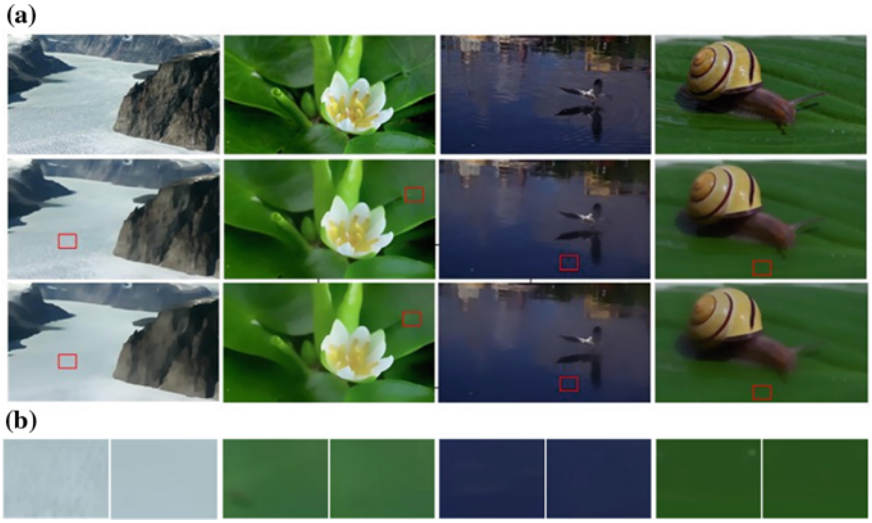

Fig. 4 Comparison between our method and the RTV. The original frames are shown in the *upper row* of (**a**). Corresponding results of RTV and our method are as shown in the *middle row* and the *bottom row*, respectively. Detail comparisons with blow-up resolutions are shown in (**b**)

In addition, we also implemented the existing RTV method for video smoothing and compared it with our proposed method. Figure 4 shows the videos smoothed using RTV and our method, respectively. For detail comparison, we selected one frame with dense textures to compare our method with RTV. As shown in Fig. 4a, the images in the top row are selected from the original videos. The middle row and the bottom row which are in Fig. 4a are the smoothed results with RTV and our method, respectively. From the visual comparisons, we can see that our smoothed results obviously are much better than the RTV results, especially for the regions marked with red boxes. From the blow-up results shown in Fig. 4b, we can even clearly observe that our method outperforms the existing RTV method, in

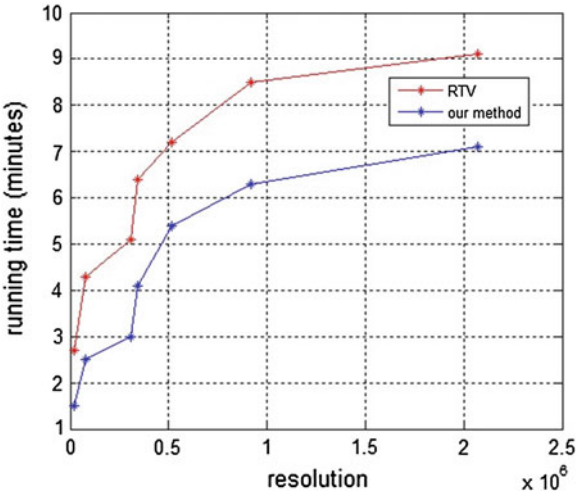

Fig. 5 Average running time comparison between RTV and our method for video with different resolutions

terms of salient structure preserving and fine detail smoothing, which points out that the combination of RTV and optical flow matching is effective in video smoothing.

On the other hand, we also performed a running-time statistics to compare our method and the RTV. In our experiments, we applied RTV and our method to 20 videos, which are about 10 s (250 frames). Each of them is up-sampled or down-sampled to be different resolutions, and we collected the average running time of RTV and our method applied on them. Figure 5 plots the average running time comparison between RTV and our method. The blue line in Fig. 5 indicates the average running time of RTV on the datasets, while the red line in Fig. 5 is ours. From the results shown in Fig. 5, we can see that our method generally faster than the RTV method, because the optical flow provides accurate correspondences and makes the optimization quickly converged. With the additional constraint of optical flow, our method turns out efficient by reducing iteration times.

5 Conclusion

In this paper, we present a novel model to handle the video texture smoothing based on RTV and optical flow matching. We contribute mainly in the following two aspects. Firstly, we set up a new probability model combined RTV with optical flow. Based on this probability model, we can easily realize the inner structure of the problem to be solved, and therefore bring up the objective function to be optimized. Secondly, we come up with an optimizing scheme to transform the original nonlinear problem to a set of subproblems that are much easier and faster to be solved. Experiments show that our method outperforms the existing

RTV method, in terms of both spatial–temporal consistency and efficiency. In the future, we plan to implement our method in GPU to achieve a real-time performance.

Acknowledgments The authors would like to thank our anonymous reviewers for their valuable comments. This work was supported in part by grants from National Natural Science Foundation of China (No. 61303101, 61170326, 61170077), the Natural Science Foundation of Guangdong Province, China(No. S2012040008028, S2013010012555), the Shenzhen Research Foundation for Basic Research, China (No. JCYJ20120613170718514, JCYJ20130326112201234, JC201005250052A, JC20130325014346, JCYJ20130329102051856, ZD201010250104A), the Shenzhen Peacock Plan (No. KQCX20130621101205783), the Start-up Research Foundation of Shenzhen University (No. 2012-801, 2013-000009), and Shenzhen Nanshan District entrepreneurship research (308298210022).

References

1. Meyer Y (2001) Oscillating patterns in image processing and nonlinear evolution equations: the fifteenth Dean Jacqueline B. Lewis memorial lectures, vol 22. American Mathematical Society
2. Yin W, Goldfarb D, Osher S (2005) Image cartoontexture decomposition and feature selection using the total variation regularized l1 functional. In: VLSM, pp 73–84
3. Xu L, Yan Q, Xia Y, Jia J (2012) ACM transactions on graphics (TOG)
4. Rudin L, Osher S, Fatemi E (1992) Nonlinear total variation based noise removal algorithms. Phys D 60(1–4):259–268
5. Aujol J-F, Gilboa G, Chan TF, Osher S (2006) Structure-texture image decomposition—modeling, algorithms, and parameter selection. Int J Comput Vis 67(1):111–136
6. Farbman Z, Fattal R, Lischinski D, Szeliski R (2008) Edge-preserving decompositions for multi-scale tone and detail manipulation. ACM Trans Graph 27:3
7. Xu L, Lu C, Xu Y, Jia J (2011) Image smoothing via L0 gradient minimization. ACM Trans Graph 30:6
8. Kass M, Solomon J (2010) Smoothed local histogram filters. ACM Trans Graph 29:4
9. Horn BKP, Schunck BG (1981) Determining optical flow. J Artif Intell 17:1852203
10. Nagel HH (1983) Displacement vectors derived from second-order intensity variations in image sequences. J Comput Vis, Graph Image Process 21(1):852117
11. Ghosal S, Vanek P (1996) A fast scalable algorithm for discontinuous optical flow estimation. J IEEE Trans Pattern Anal Mach Intell 18(2):1812194
12. Lucas B, Kanade T (1981) An iterative image registration technique with an application to stereo vision. In: Proceedings of DARPA Image Understanding Workshop, p 1212130

Springer

http://www.springer.com/978-3-642-54926-7

Practical Applications of Intelligent Systems
Proceedings of the Eighth International Conference on
Intelligent Systems and Knowledge Engineering,
Shenzhen, China, Nov 2013 (ISKE 2013)
Wen, Z.; Li, T. (Eds.)

2014, XVII, 1176 p. 550 illus., Softcover
ISBN: 978-3-642-54926-7