Dynamic Texture Segmentation Using Fourier Transform

Jianghong Li

School of Power and Energy, Northwestern Polytechnical University

Xi'an 710072, China

Tel: 86-029-8846-0464   E-mail: jhli@nwpu.edu.cn

Liang Chen

School of Power and Energy, Northwestern Polytechnical University

Xi'an, Shaan'xi, China 710072

Tel: 86-029-8846-0464   E-mail: xbgydxxs@163.com

Yuanhu Cai

School of Power and Energy, Northwestern Polytechnical University

Xi'an710072, China

Tel: 86-029-8846-0464   E-mail: caiyh@nwpu.edu.cn

Abstract

Dynamic textures are temporally continuous and infinitely varying sequences of images with certain spatial and temporal stationarity properties and have many potential applications such as computer graphics, computer vision, animation, and computer games. These applications usually require dynamic texture segmentation that is not an easy task especially when the background is cluttered and textured. In this paper, we analyze the spatio-temporal slices of dynamic texture using Fourier analysis and then we propose a simple and efficient approach to dynamic texture segmentation using 3D Fourier transform. Experiments on several dynamic texture sequences show the effectiveness of the proposed approach.

Keywords: Dynamic texture segmentation, Fourier transform

1. Introduction

Dynamic textures are temporally continuous and infinitely varying sequences of images with certain spatial and temporal stationarity properties. Typical examples of dynamic textures include sea-waves, moving clouds, smoke, foliage, fire, walking crowd, and highway traffic. They are common in natural scenes and each of them possesses an inherent dynamics and repetitive pattern. Dynamic textures can easily be perceived by human due to their simplicity and coherence. However, they are difficult to be analyzed as the underlying motion and pattern are often complex and stochastic. The importance of analyzing dynamic textures lies in their relevance to the research areas of computer vision, video processing, and computer graphics. Dynamic textures have many potential applications such as computer graphics, computer vision, animation, computer games, and automated surveillance. In recent years, lots of research efforts have been spent on dynamic texture analysis and synthesis especially in the areas of computer vision and computer graphics. Dynamic texture analysis is to exploit the spatial and temporal properties to characterize or represent dynamic textures for analysis and recognition. While the goal of dynamic texture synthesis is to generate or reconstruct a new video sequence which is similar to the original sequence of dynamic texture without any noticeable artifacts. Both dynamic texture analysis and synthesis usually require dynamic texture segmentation that extract the regions of dynamic textures from a video before performing analysis and synthesis. The spatial extent of dynamic textures (e.g., smoke, fire, flow water) can vary over time and they can also be partially transparent. Therefore, segmenting dynamic textures from a complex background is not an easy task especially when the background is cluttered and textured. This paper considers developing a simple and efficient approach to dynamic texture

segmentation. The rest of the paper is organized as follows. The related work is reviewed in Section 2. The proposed approach is presented in Section 3, and experimental results are demonstrated in Section 4. Final conclusions are summarized in Section 5.

2. Related work

Numerous techniques have been proposed in the literature for dynamic texture segmentation. These techniques can be classified into three major categories: (1) model-based, (2) motion-based, and (3) feature-based techniques. Model-based techniques (M. O. Szummer, 1995)(B. U. Toreyin and A. E. Cetin, 2007)(G. Doretto, A. Chiuso, Y. N. Wu, and S. Soatto, 2003)(A. B. Chan and N. Vasconcelos, 2005)(B. Abraham, O. I. Camps, and M. Sznaier, 2005)(B. Ghanem and N. Ahuja, 2007) are to model the spatio-temporal interdependence among images. Szummer (M. O. Szummer, 1995) proposed the spatio-temporal auto-regressive (STAR) model. In (B. U. Toreyin and A. E. Cetin, 2007)(B. Ghanem and N. Ahuja, 2008), hidden Markov model is adopted to model the dynamic textures. Doretto et al. (G. Doretto, A. Chiuso, Y. N. Wu, and S. Soatto, 2003) derived a stable linear dynamical system model for dynamic textures. In this model, consecutive frames of a sequence are linearly related and viewed as the responses of the linear dynamical system to random noise input. This model has been applied to dynamic texture segmentation(G. Doretto, D. Cremers, P. Favaro, and S. Soatto, 2003). In (A. B. Chan and N. Vasconcelos., 2005), linear dynamical system has been extended to accommodate a mixture of dynamic textures that has been utilized for crowd and traffic segmentation. Abraham et al. (B. Abraham, O. I. Camps, and M. Sznaier, 2005) proposed to identify the underlying linear dynamic system using a set of Fourier descriptors of the image frames rather than the original sequence. All these techniques model the intensity values of a dynamic texture as a stable and linear ARMA process that leads to significant computational expense because the model directly to pixel intensities without mitigating spatial redundancy. To overcome the limitations of these techniques, Ghanem et al. (B. Ghanem and N. Ahuja, 2007) proposed a model that relates texture dynamics to the variation of the Fourier phase, which captures the relationships among the motions of all the pixels within the texture as well as the appearance of texture. Motion-based techniques (T. Amiaz, S. Fazekas, D. Chetverikov, and N. Kiryati, 2007)(S. Fazekas and D. Chetverikov, 2005)(S. Fazekas, T. Amiaz, D. Chetverikov, and N. Kiryati, 2009)(R. Peteri and D. Chetverikov, 2005)(R. Vidal and A. Ravichandran, 2005) are based on motion estimation algorithms in which frame-to-frame motion field is estimated. Motion estimation has been extensively studied and computationally efficient algorithms have been developed. The popular motion estimation algorithm used for dynamic textures is optical flow algorithm. Optical flow is usually based on the assumption of brightness constancy, i.e., the brightness of an object is constant from frame to frame. In addition, most of optical flow algorithms are often suited to estimate local and smooth motion fields. However, the spatial extent of dynamic textures can vary over time rapidly and is often not smooth. These characteristics make it difficult for optical flow algorithm to estimate motion effectively. Feature-based techniques (K. Otsuka, T. Horikoshi, S. Suzuki, and M. Fujii, 1998)(G. Zhao and M. Pietiknen, 2006)(A. Rahman and M. Murshed, 2007) use image features to characteristics of dynamic textures. In (K. Otsuka, T. Horikoshi, S. Suzuki, and M. Fujii, 1998), spatiotemporal motion trajectory is utilized for feature extraction for dynamic textures. Zhao et al. (G. Zhao and M. Pietiknen, 2006) proposed to use local binary pattern, which has been used for 2D textures, to describe dynamic textures for recognition tasks. In (A. Rahman and M. Murshed, 2007), motion co-occurrence matrix is employed to characterize dynamic textures and a segmentation method based on spatial and temporal motion co-occurrence statistics is presented.

3. Dynamic texture segmentation

In this paper, we consider developing an efficient approach to dynamic texture segmentation based on Fourier analysis. The propose approach is motivated by the work presented in (B. Ghanem and N. Ahuja, 2007), which forms an efficient spatio-temporal model of both the appearance and global dynamics of a dynamic texture using Fourier phase. Before describing the proposed approach, we would like to review the Fourier analysis for dynamic textures.

3.1 Fourier analysis

Fourier analysis has been widely used in image processing as it has several useful properties. Fourier analysis is robust against perturbations that often appear in images, for example, illumination changes and additive noises. The frequency spectrum of an image can be calculated by using the fast Fourier transform (FFT) algorithm that is practical and computationally efficient. In (B. Ghanem and N. Ahuja, 2007), the phase content of a dynamic texture has been demonstrated to be useful for representing its appearance and temporal variations based on the assumption that symmetric Z-transform factors seldom occur in practice. In addition, the temporal variations of phase contents are empirically shown to capture most of dynamical characteristics of a dynamic texture. Motivated by this, we consider using the phase of Fourier analysis for dynamic texture segmentation. Let us choose the sequence "644ce10.avi" from (R. Peteri, S. Fazekas, and M. Huiskes, 2006) as an example. Frame 30 of this sequence is depicted in Figure 1(a). Such sequence mainly contains water waves that can be considered as a dynamic texture and a floating duck

---

Algorithm 1 The proposed approach to dynamic texture segmentation

---

Require: An input image sequence I(x, y, t) **Ensure:** The segmentation mask for each frame

---

1.   *for t from 1 to T do*

2.   *Downsample the current frame of I(x, y, t);*

3.   *Smooth the current frame using 2D Gaussian filter;*

4.   *end for*

5.   *Perform 3D FFT for the whole sequence I(x, y, t) using (1);*

6.   *Calculate the phase spectrum using the real and imaginary parts of 3D DFT;*

7.   *Calculate the reconstructed sequence Î(x, y, t) using (2);*

8.   *for t from 1 to T do*

9.   *Smooth the reconstructed frame of ˆI(x, y, t) using the avergaing filter;*

10.   *Compute the mean value ˆμ(t) of the current frame;*

11.   *Convert the current frame into binary image using ˆμ(t) as the threshold;*

12.   *Perform morphological processing, i.e., filling and closing, to obtain binary mask for the current frame;*

13.   *Up-sample the binary mask;*

14. end for

---

in a lake. For simplicity, we take a slice of this sequence for analysis. The vertical white line shown in Figure 1(a) is the location of X-T slice, which covers pixels from water waves, the duck, and the background. The corresponding X-T slice image is depicted in Figure 1(b). We take 2D Fourier transform of this slice image using 2D FFT. The log power spectrum and phase spectrum obtained from the Fourier transform are shown in Figure 1(c) and Figure 1(d), respectively. We can see that the phase contains more significant information on the structure in the image than the magnitude. We now consider taking the magnitude and phase separately to reconstruct the slice image using 2D inverse FFT. Figure 1(e) and Figure 1(f) demonstrate the reconstructed slice images using the magnitude and phase, respectively. It is interesting to see that the reconstructed slice image from the phase carries essential information about the water waves, i.e., a kind of dynamic textures. However, the reconstructed slice image from the magnitude provides little information.

3.2 Proposed approach

In this section, we present the proposed approach to dynamic texture segmentation. According to the discussion in Section 3.1, we consider using 3D discrete Fourier transform to process the dynamic textures.

We now describe the details of the proposed approach. Given a gray-level image sequence I(x, y, t) containing dynamic textures where (x, y) is the location of each pixel and t is the frame index. We perform the 3D discrete Fourier transform (DFT) of the input sequence I(x, y, t):

$$F(I(x,y,t)) = \sum_{x=0}^{X-1}(\omega_X^{k_1 x}\sum_{y=0}^{Y-1}(\omega_Y^{k_2 y}\sum_{t=0}^{T-1}\omega_T^{k_2 t}I(x,y,t)))\tag{1}$$

where X and Y are the width and height of each frame, respectively. T is the total number of frames. In (1), $\omega X = \exp(-2\pi i/X)$ and $k1 = 0, 1, \cdots, X-1$. After performing 3D DFT of I(x, y, t), the phase spectrum P[F(I(x, y, t))] can be obtained by using the real and imaginary parts of F(I(x, y, t)). We then compute the reconstructed sequence by performing 3D inverse FFT on the phase spectrum only, i.e.,

$$\hat{I}(x,y,t)) = \left| F^{-1}\left\{ e^{iP[F(I(x,y,t))]} \right\} \right|^2 \tag{2}$$

Where Î(x, y, t) represents the reconstructed sequence.

Each frame of the reconstructed sequence is smoothed by using an averaging filter. Here we select the circular averaging filter with the radius R. The resulting smoothed image is subsequently converted to a binary image by using thresholding. In our approach, the mean value of the image ˆμ(t) is chosen to be the threshold. According to our experimental results, such thresholding technique works quite well. After the binary image is obtained, the morphological processing, i.e., filling and closing, is performed as post-processing step to obtain the segmentation mask for each frame. In order to reduce the computational cost, each frame of the input sequence can be down-sampled and smoothed by a Gaussian filter before performing 3D DFT. In addition, down-sampling the image followed by

smoothing can reduce the noise in the image. If such preprocessing step is employed, the segmentation mask is required to be up-sampled. The proposed approach is summarized in Algorithm 1. It is seen that the proposed approach is quite simple and the most computational expensive step is 3D DFT, which can be computed using efficient 3D FFT.

4. Experimental results

In this section, we test our proposed approach to dynamic texture segmentation for dynamic texture sequences that are available from (R. Peteri, S. Fazekas, and M. Huiskes, 2006). The details of these sequences are described in Table 1. The image size of each sequence is 352×288. The format of each sequence is MPEG-4 Divx. In all of our experiments, the down-sampling and up-sampling factors are set to 2. The radius of the circular averaging filter is set to 9 for each frame of the reconstructed sequence. The proposed approach is implemented in MATLAB and all the experiments are conducted on a Pentium 4 laptop with 2 GB of RAM.

Figure 2 demonstrates the segmentation results achieved by using our proposed approach to dynamic texture segmentation. The image on the left side of Figure 2(a), (b), (c), (d), (e), (f), (g), (h), (i), and (j) shows the original image, while the image on the right side shows the original image overlaid with the corresponding segmentation mask. We can see that the proposed approach can achieve reasonably good segmentation for dynamic textures especially when the camera is static or the camera motion is small.

5. Conclusions

In this paper, the Fourier analysis for spatio-temporal slices of dynamic texture sequence has been conducted and indicates that the phase spectrum is important for dynamic texture segmentation. Such analysis motivates us to have proposed a simple and efficient approach to dynamic texture segmentation using 3D Fourier transform, which can be computed using FFT. We have applied the proposed approach to a variety of dynamic texture sequences and experimental results show that it is effective for dynamic texture segmentation.

References

A. B. Chan and N. Vasconcelos. (2005). Mixtures of dynamic textures. *Tenth IEEE International Conference on Computer Vision*, 2005., 1, 2005.

A. Rahman and M. Murshed. (2007). Segmentation of dynamic textures. *Computer and Information Technology*, 2008. ICCIT 2008, pages 1–6, 2007.

B. Abraham, O. I. Camps, and M. Sznaier. (2005).Dynamic texture with fourier descriptors. Proc. of the 4th Int. Workshop on Texture Analysis ans Synthesis, pages 53–58, 2005. Proc. Of the 4th Int. Workshop on Texture Analysis ans Synthesis.

B. Ghanem and N. Ahuja. (2007). Phase based modelling of dynamic textures. *IEEE 11th International Conference on Computer Vision*, 2007.

B. Ghanem and N. Ahuja. (2008). Extracting a fluid dynamic texture and the background from video. *IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pages 1–8, 2008.

B. U. Toreyin and A. E. Cetin. (2007). Hmm based method for dynamic texture detection. IEEE 15th Signal Processing and Communications Applications, 2007, pages 1–5, 2007.

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

G. Doretto, D. Cremers, P. Favaro, and S. Soatto. (2003). Dynamic texture segmentation. *IEEE International Conference on Computer Vision*, 2:1236–42, 2003. IEEE International Conference on Computer Vision.

G. Zhao and M. Pietiknen. (2006). Dynamic texture recognition using volume local binary patterns. Proc. ECCV 2006 Workshop on Dynamical Vision, pages 165–177, 2006.

K. Otsuka, T. Horikoshi, S. Suzuki, and M. Fujii. (1998). Feature extraction of temporal texture based on spatio-temporal motion trajectory. *International Conference on Pattern Recognition*, 1998, 2, 1998.

M. O. Szummer. (1995). Temporal texture modeling. PhD thesis, Massachusetts Institute of Technology, 1995.

R. Peteri and D. Chetverikov. (2005). Dynamic texture recognition using normal flow and texture regularity. Proc. Iberian Conference on Pattern Recognition and Image Analysis, pages 223–230, 2005.

R. Peteri, S. Fazekas, and M. Huiskes. (2006). Dyntex: A comprehensive database of dynamic textures. http://www.cwi.nl/projects/dyntex, 2006.

R. Vidal and A. Ravichandran. (2005). Optical flow estimation and segmentation of multiple moving dynamic textures. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, 2.

S. Fazekas and D. Chetverikov. (2005). Normal versus complete flow in dynamic texture recognition: a comparative

study. Int. Workshop on Texture Analysis and Synthesis, 2005. *Int. Workshop on Texture Analysis and Synthesis*.

S. Fazekas, T. Amiaz, D. Chetverikov, and N. Kiryati. (2009). Dynamic texture detection based on motion analysis. *International Journal of Computer Vision*, 82(1):48–63, 2009.

T. Amiaz, S. Fazekas, D. Chetverikov, and N. Kiryati. (2007). Detecting regions of dynamic texture. Lecture Notes in Computer Science, 4485:848, 2007.

Table 1. Dynamic texture sequences used in our experiments

No.	Sequence	Description
1	"6ammb00.avi"	Small candles with flames
2	"73v192u.avi"	Smoke, fire, and steam
3	"644ce10.avi"	Duck floating on water
4	"645c620.avi"	Highway traffic
5	"648ea10.avi"	Smoke from ventilation system
6	"6481h10.avi"	Water falling over pebbles
7	"6481n10.avi"	A small winding creek
8	"6483b10.avi"	The close up of "6481n10.avi"
9	"6482f10.avi"	A small lake with water ripples
10	"6488610.avi"	Fountain and water ripples

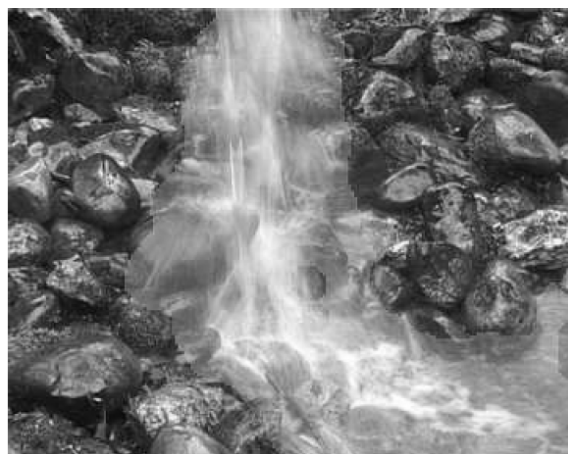(a) "6ammb00.avi"

(b) "73v192u.avi"

(c) "644ce10.avi"

(d) "645c620.avi"

(e) "648ea10.avi"

(f) "6481h10.avi"

(g) "6481n10.avi"

(h) "6483b10.avi"

(i) "6482f10.avi"

(j) "6488610.avi"

Figure 2. Segmentation results obtained by using the proposed approach for some dynamic texture sequences (R. Peteri, S. Fazekas, and M. Huiskes, 2006). The images on the left and right sides of (a), (b), (c), (d), (e), (f), (g), (h), (i), and (j) depict the original image and the corresponding mask with red color, respectively. This figure is best viewed in color.