PHASE PCA FOR DYNAMIC TEXTURE VIDEO COMPRESSION

Bernard Ghanem, Narendra Ahuja

Beckman Institute for Advanced Science and Technology
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign
Urbana, IL 61801, USA
{bghanem2, ahuja}@vision.ai.uiuc.edu

ABSTRACT

Temporal or dynamic textures (DT's) are video sequences that are spatially repetitive and temporally stationary. DT's are temporal analogs of the well known spatial still image texture. Examples of DT's include moving water, foliage, smoke, clouds, etc. We present a new DT model that can efficiently compress DT sequences. Our proposed method compactly represents the spatiotemporal properties of a DT by modelling its varying Fourier phase content, which can be shown to be the major determinant of both its dynamics and appearance. This is possible because this method combines both temporal and spatial properties in a compact spectral framework. Making use of the benefits inherent to working in the frequency domain, this model provides a significant improvement in DT compression, which can be used to improve the performance of MPEG-2 encoding. We will present experimental evidence that validates this method for a variety of complex sequences, while also comparing it to the most recent DT representational model that is based on modelling a DT as a linear dynamical system (LDS).

***Index Terms*—** Dynamic texture, Phase, LDS, PCA

1. INTRODUCTION

Modelling of complex motion patterns in images remains unsolved in computer vision, since it poses numerous problems especially those related to reliable motion field estimation. These problems become even more complex when considering non-rigid stochastic motions (e.g. DT's). For example, a scene of "translating" clouds conveys visually identifiable global dynamics; however, the implosion and explosion of the cloud segments during the motion result in very complicated local dynamics. So, it is evident that efficient DT compression poses a serious challenge.

Methods ([1]) relying on optical flow are convenient, since frame-to-frame estimation of the motion field has been extensively studied and computationally efficient algorithms have been developed. However, these methods only capture temporal characteristics of the DT and are prone to error especially due to noise sensitivity and motion discontinuity. In fact, motion field estimation becomes a significantly harder task due to the non-rigid nature and complex motion prevalent in DT's.

In comparison to the previous techniques, fewer spatiotemporal models have been developed for DT's. These include the pioneering work by Nelson and Polana (1992) [2], the spatio-temporal auto-regressive (STAR) by Szummer and Picard [3], and multi-resolution analysis (MRA) trees by Bar-Joseph et al. (2001) [4]. These methods impose restrictions on the textures that can be modelled or are applied directly on pixel intensities instead of more compact representations (i.e. pixel groupings), thus, precluding feasible compression. The most recent representational DT model was developed by Doretto et al. (2003) [5], in which a linear time invariant dynamical model (LDS) is derived for the DT. This model has been applied to DT compression and synthesis [5], recognition [6], and segmentation [7]. However, its modelling of the intensity values of a DT as a stable, linear ARMA (1) process leads to three main disadvantages: (i) the assumption of second-order probabilistic stationarity, which does not hold for numerous sequences (e.g. fire), (ii) the suboptimal relationship between the order of the LDS model and the extent of temporal modelling possible (i.e. an LDS of order n does not capture the most temporal variation in a DT among all models of order n), and (iii) significant computational expense, since the model is applied directly to pixel intensities without appropriately mitigating spatial redundancy.

Our method can be categorized as a spatiotemporal, image-based model that uses the Fourier phase content of the DT sequence to model both its appearance and global dynamics. The rest of this paper is organized as follows: we justify our choice of using phase in Section 2, present the details of our phase based compression model in Section 3, and provide experimental results that compare its performance to that of LDS and a standard video compression scheme, MPEG-2.

2. MOTIVATION

Following are the advantages of using the frequency domain representation that alleviate certain problems encountered in the spatial domain and motivate our proposed approach. **(1)** Spatially global features are captured locally in the frequency domain, since the change of the amplitude or phase of a certain frequency results in a global spatial variation. This makes frequency space modelling more appropriate for modelling global patterns such as those associated with DT appearance and dynamics. **(2)** Computational complexity can be reduced by exploiting the inherent conjugate symmetry of the Fourier transform and the concentration of spectral image energy at low frequencies. **(3)** Furthermore, computationally efficient algorithms and specialized hardware are available for the computation of the Fourier transform (e.g. FFT).

In what follows, we justify why the phase content of a DT is a useful dual representation of its appearance and temporal variations, and leads to a compact spatiotemporal model. **(1)** In [8], Hayes proved that it is possible to reconstruct multi-dimensional signals from their phase content alone, provided that these signals do not have symmetric factors in their Z-transforms. In fact, if a hybrid image is constructed from the phase spectrum of a given image and the amplitude spectrum of any other, we use the iterative algorithm, described in [9], to reconstruct the original image from the hybrid image. This process is called phase-only reconstruction. In each iteration, two stages of processing are used: a 2D FFT (twice the size of the input image) followed by a 2D inverse FFT (same size as the input image). Figure 1 shows an example of this algorithm applied to ocean and fire images. In the rest of this paper, we assume that DT sequences enjoy this phase-only reconstruction property. This assumption is justified, since symmetric Z-transform factors seldom occur in practice.

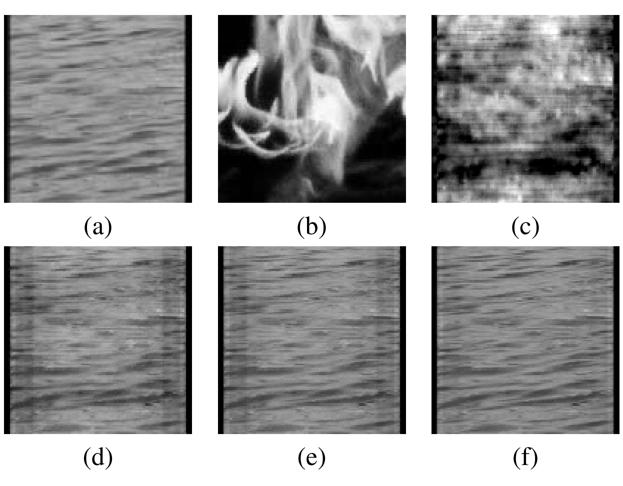


Fig. 1. (a) is the phase spectrum image. (b) is the amplitude spectrum image. (c) is the hybrid image. (d), (e), and (f) are the images after 50, 100, and 250 iterations respectively.

(2) Complex stochastic motion, which characterizes a DT, leads to complex stochastic variations in its phase content. We empirically show that the temporal variations of phase values do indeed capture most of the DT's dynamical characteristics and hence its global motion. Figure 2 shows that many more principal components are required to represent 80% of the variation in DT phase than to represent the same amount in DT amplitude. Hence, we conclude that it is relevant to represent only the phase spectrum of DT's for the purpose of compression.

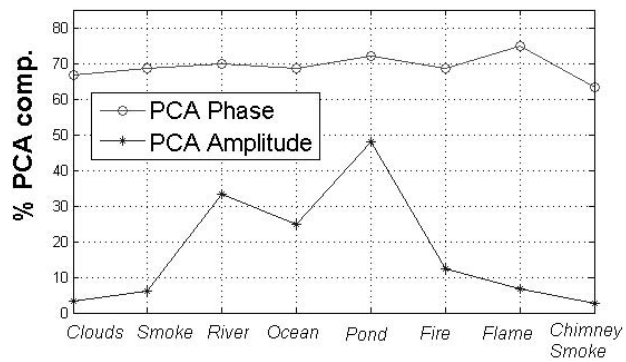


Fig. 2. PCA components for DT phase and amplitude.

3. PHASE PCA COMPRESSION MODEL

For DT compression, we perform PCA on the DT feature vectors, which are the vectorized half spectra of DT phase for the frames in the DT sequence. Only half the phase spectrum is required due to the conjugate symmetry property of the FFT. Here, we note that the DT frames are preprocessed with an appropriately sized Hanning window to mitigate spectral leakage. We will use the notation in (1) and (2) to denote the PCA basis (**A**), the feature vectors ($\vec{\Phi}$), their projections in the PCA space ($\vec{x}$), and the mean feature vector ($\vec{\Phi}_m$). For an image of size MxN, the length of the feature vector is K = $\frac{MN}{2}$, so that the size of **A** is KxL, where L is the number of principal components that have been selected to represent the data. For complete representation, L = F, which is the number of frames in the DT sequence.

$$\vec{\Phi} = A\vec{x} + \vec{\Phi}_m \quad (1)$$

$$\vec{x} = A^T(\vec{\Phi} - \vec{\Phi}_m) = [x_1 x_2 \cdots x_L]^T \quad (2)$$

Due to the symmetry of DT phase, considerable compression is possible for each individual frame. We note here that additional compression can be achieved by neglecting frequencies with low energy, mainly in the high frequency bands and by using the dimension reduction option inherent to PCA. Below, we first present the compression rates that can be achieved by both Phase PCA and LDS in terms of the number of principal components used for DT representation. Then, we describe how the compression performance of our method can be enhanced by forming a more compact PCA space.

Basic Phase PCA (BPP) Compression: In this section, we will present the overall compression rate that can be achieved by reducing the dimensionality of the phase PCA space. So, with this generic layout, we can compute the expected overall compression rate (R_{comp}) for an arbitrary DT sequence as in (3). In fact, since L $\leq$ F and $\frac{MN}{2} \gg$ F, then the removal of a PCA component will lead to significant data compression.

$$\begin{aligned} R_{\text{comp}} &= 1 - \frac{\text{size}(\mathbf{A}) + \text{size}(\vec{\Phi}_m) + \text{\# of PCA coefficients}}{MNF} \\ &= 1 - \left[\frac{L+1}{2F} + \frac{L}{MN}\right] \approx 1 - \frac{L+1}{2F} \quad (L \ll MN) \end{aligned} \tag{3}$$

The main factor dictating the extent of the data compression is $\frac{L}{F}$, the fraction of the PCA components used in the representation. Also, note that even with a complete representation ($L = F$), the compression rate is about 50%. This is due to the fact that only half the phase spectrum is used to represent the DT, so the amplitude spectrum must be initially determined by the iterative process mentioned in the context of phase-only reconstruction.

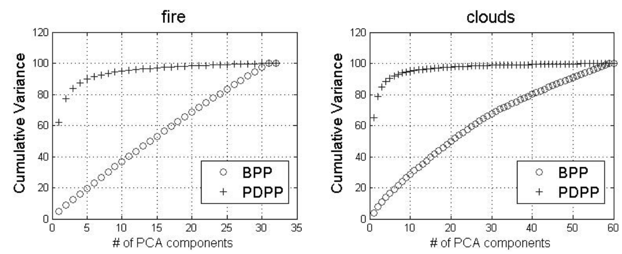
Using LDS, we require two matrices (**A** and **C**) and the initial state $\vec{x}_0$ in order to reconstruct the DT. The dimension of **A** is L'xL' and that of **C** is $2K$xL', where L' represents the order of the LDS and $K = \frac{MN}{2}$ as defined before. So, the overall compression rate in the case of a DT with F frames is estimated as in (4). Note that under the same compression rate, the LDS method requires approximately half the number of principal components needed by BPP.

$$\begin{aligned} R_{\text{comp}} &= 1 - \frac{\text{size}(\mathbf{A}) + \text{size}(\mathbf{C}) + \text{size}(\vec{x}_0)}{MNF} \\ &= 1 - \frac{L'^2 + 2KL' + L'}{MNF} \approx 1 - \frac{L'}{F} \end{aligned} \tag{4}$$

Principal Difference Phase PCA (PDPP) Compression: In this section, we describe an equivalent DT phase spectrum, which forms a framework that allows for further DT compression. PDPP represents the phase changes in terms of the principal angle of the difference between phase spectra of consecutive frames. We transform the BPP phase into PDPP format as follows. Each extracted phase spectrum is vectorized (Φ_i^r, Φ_i^{r+1}) and replaced by the sum of the previous phase spectrum and the principal angle of their difference, between Φ_i^{r+1} and Φ_i^r $\forall r = 1, \ldots, F$ $\forall i = 1, \ldots, M'$ as illustrated in Equation 5. In fact, this transformation expands the domain of the original BPP space by 2π in each dimension. Hence, the PDPP space can be spanned by fewer principal components, giving rise to a more compact spatiotemporal model.

$$\begin{aligned} &\Phi_i^{r+1} \longleftarrow \Phi_i^r + \Gamma(\Phi_i^{r+1} - \Phi_i^r) \\ &\Gamma(x) = x + 2\pi k \ \in \ ]-\pi, \pi], \text{ for some } k \in \mathbb{Z} \end{aligned} \tag{5}$$

In Figure 3, we show that for the same number of PCA components (i.e. compression rate), PDPP can represent significantly more variation in DT phase than the BPP method described previously.



(a) (b)

Fig. 3. BPP vs. PDPP for two DT's

4. EXPERIMENTAL RESULTS

In this section, we present experimental results that validate the significance of our proposed method for DT compression and compare its performance to that of LDS. Figure 4 ((a)-(c)) compares the performance of LDS, BPP, and PDPP over a range of compression rates, that are proportional to the number of principal components used. We note here that the compression rate is computed from the number of principal components required by LDS. It is evident from these plots that BPP, in general, outperforms the LDS compression scheme, mainly due to the fact that only half the phase spectrum is modelled. Furthermore, PDPP renders a significant improvement over BPP even at very low compression rates. In (d), we show the temporal performance of each compression scheme at a compression rate of 63%. We notice that both BPP and PDPP tend to oscillate about a steady PSNR value, while LDS performance decreases with time. This is due to the fact that LDS produces a DT frame as a linear combination of the L' chosen principal components, which are computed from L' frames of the original DT and not all of them (i.e. F).

Next, we compare PDPP compression to that of the MPEG-2 standard, as portrayed in Figure 5. Here, we define the compression rate for each case as the ratio of the size of the MPEG-2 video to that of the original, uncompressed video.

From the above plots, we see that our method outperforms MPEG-2 in all four DT's. This improvement is primarily due to the more compact representation of the temporal characteristics of the DT inherent to PDPP. Since MPEG-2 requires computation of motion fields and these estimates based on

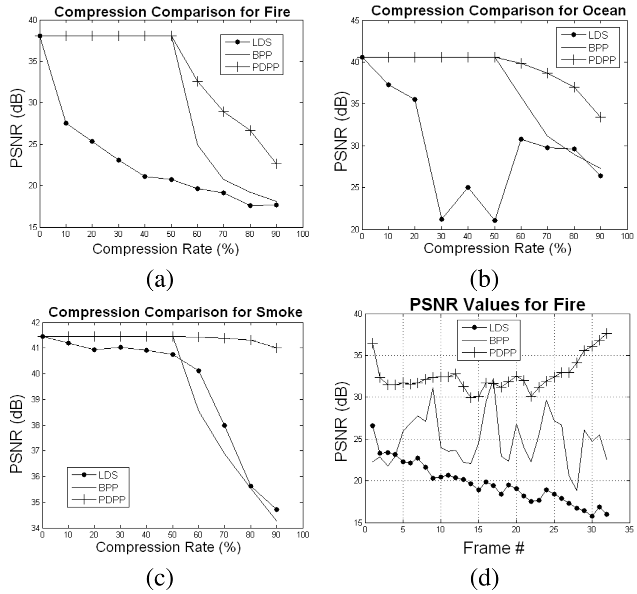


Fig. 4. LDS vs. BPP vs. PDPP in terms of compression rate

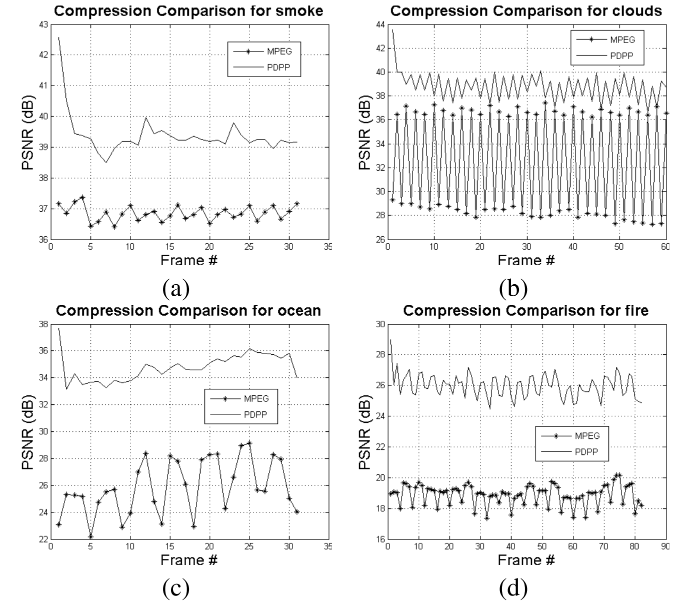


Fig. 5. MPEG-2 vs. PDPP at a compression rate of approximately 60% for (a)-(d)

optical flow algorithms tend to degrade with the complexity of the motion and the moving objects, MPEG-2 does not perform as well for the stochastic motion of non-rigid particle objects prevalent in DT's.

5. CONCLUSION AND FUTURE WORK

In this paper, we proposed a novel compression model for DT's, which represents both appearance and temporal information based on DT phase content. This model was shown to outperform the most recently used DT model in the literature. Moreover, we compared our model with MPEG-2 and showed that more compression is possible, when a more compact representation of the temporal properties of the DT is available. In the future, we would like to extend this model to incorporate only high energy frequencies and examine how information theoretic coding might improve its compression performance.

6. REFERENCES

[1] P. Bouthemy and R. Fablet., "Motion characterization from temporal cooccurrences of local motion-based measures for video indexing," in *Proc. of ICPR*, 1998, vol. 1, pp. 905–908.

[2] R.C. Nelson and R. Polana, "Qualitative recognition of motion using temporal texture," in *Proc. of ICPR*, 1992, pp. 56–78.

[3] M. Szummer and R. W. Picard, "Temporal texture modeling," in *Proc. of ICIP*, 1996, vol. 3.

[4] Z. Bar-Joseph, R. El-Yaniv, D. Lischinski, and M. Werman, "Texture mixing and texture movie synthesis using statistical learning," *IEEE Trans. on Visualization and Computer Graphics*, pp. 120–135, 2001.

[5] S. Soatto, G. Doretto, and Y. N. Wu, "Dynamic textures," *International Journal of Computer Vision*, vol. 51, pp. 91–109, 2003.

[6] P. Saisan, G. Doretto, Y. N. Wu, and S. Soatto, "Dynamic texture recognition," in *Proc. of ICPR*, 2001, vol. 2, pp. 58–63.

[7] G. Doretto, D. Cremers, P. Favaro, and S. Soatto, "Dynamic texture segmentation," in *Proc. of ICCV*, 2003, vol. 2, pp. 1236–1242.

[8] M. Hayes, "The reconstruction of a multidimensional sequence from the phase or magnitude of its fourier transform," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 30, no. 2, 1982.

[9] A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," *Proceedings of the IEEE*, vol. 69, pp. 529–541, 1981.