High Dimensional versus Low Dimensional Chaos in *MPEG-7* Feature Binding for Object Classification*

Hanif Azhar and Aishy Amer

Electrical and Computer Engineering,
Concordia University, Montréal, Québec, Canada
{h_azhar,amer}@ece.concordia.ca

Abstract. We proposed [1] a feature binding method to generate a *MPEG-7* compliant feature vector, defined as *C-MP7*. Here, we study the excellence of *C-MP7* as a feature vector, using either low- or high-dimensional chaos. With high-dimensional chaos-based *C-MP7*, we find, 1) the accuracy in SVM classifier improves 10% to 20%, for all classes of video objects over *MPEG-7*, 2) larger binary class separation among video objects in different classes, 3) *vehicle* objects are clustered well, which leads to above 99% accuracy for only *vehicles* against other objects in SVM, and 4) drifts in chaotic attractors allow the *C-MP7* to include subtle variations in coefficients for video objects.

Keywords: Chaos, *MPEG-7*, Video, Object, Feature.

1 Introduction

To improve non-MPEG-7 feature-based simple video object classification (e.g., [2]), some surveillance works (e.g., [3]) have reported the use of individual MPEG-7 visual descriptors. Each MPEG-7 visual descriptor has multiple feature coefficients. Each visual descriptor defines a low-level feature, and when combined, multiple visual descriptors usually constitute high-dimensional feature vector. Such feature vector can be vital to represent video objects. However, the correlation among feature coefficients in different visual descriptors is not defined, and the descriptors are not prune to feature redundancy [4]. *MPEG-7* feature vector needs to be refined for media content in specific applications.

In surveillance, video objects can be challenging, i.e., different scales, lighting conditions, resolutions, and orientations, posture variations for same object in successive frames, existence of parts of backgrounds or other objects, lost or missing parts of video objects in some frames. Discriminative training patterns for the same video object, are not always available from video analysis. Earlier in [1], we investigate scopes of chaotic series in *MPEG-7* feature vector, propose a new feature vector *C-MP7*, and report improvements in cross-validation accuracy

* This work was supported, in part, by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

with similar video objects from successive frames. Our work is inspired by the existence of chaotic series in the neuronal response of human brain activities [5].

Here, we present new analysis which validates the excellence of *C-MP7* over *MPEG-7*, for diverse video objects. We study, *C-MP7* with both high-dimensional (*h-d*, here after) and low-dimensional (*l-d*, here after) chaos for classification of diverse video objects from public and local databases. With *h-d* chaos, we find, 1) when compared with *MPEG-7*, the accuracy in Support Vector Machine (SVM) classifier improves 10% to 20% for all classes of video objects, with *C-MP7*, but reduces with *C-MP7* (*l-d* chaos), 2) multi-class Fisher-criteria shows increased binary class separation, with *C-MP7*, 3) *vehicle* objects are clustered well with *C-MP7*, which leads to above 99% accuracy for only *vehicles* against other objects in SVM, and 4) drifts in attractors due to transient allow *C-MP7* to include subtle variations in coefficients for video objects in a class. A higher variance of *C-MP7* with *h-d* chaos over that with *l-d* chaos, manifests the subtle variations.

In the rest of this paper, Sec. 2 describes the background on brain activity and chaos theory, Sec. 3 recalls the core simulation approach [1]. Latest analysis on *C-MP7* is presented in Sec. 4, and the conclusion is reported in Sec. 5.

2 Brain and Chaos

There has been a established link between the analysis of temporal brain oscillatory signals and chaotic dynamics from the EEG signals [5] in mass neuronal activity of the brain. The visual features of any real world object get projected in the retina and then is conveyed to the primary visual cortex, called V1, and subsequently these features cause neurons there to behave as a combination of excitory and inhibitory nodes. From V1, the nueronal response of visual features is sent to several neighboring higher visual areas in the brain. Each of these areas sends its output to several others. Less is known about how the visual features are processed after the V1, i.e., visual area 2, or the third area, or the middle temporal area [6]. However, the brain is a non-linear system par excellence and chaotic dynamics can be used to explain the ability of the brain to generate and process information [5]. Neuronal excitations in the brain are heavily dependent on chaotic attractors [7]. The chaotic attractor is a non-linear geometrical structure generated by a chaotic series. A chaotic series can be generated from non-linear one dimensional finite-difference or ordinary differential equation. Interesting phenomena can arise in the local stability analysis of a simple finite-difference equation. Quadratic map, also known as Logistic map, is such a nonlinear equation,

$$x_{t+1} = Rx_t(1 - x_t), \quad (1)$$

where $0 < t \leq \infty$. Different range of R , reveals different behavior by Eq. 1. With $R = 4$, the solution to Eq. 1 can oscillate in aperiodic manner. This oscillation shows neither exponential growth or decay, nor a steady state. This

behavior is chaos, and Eq. 1 is a chaotic series. A fixed point of a function $f(x_t)$ is a value x_t^* that satisfies $x_t^* = f(x_t^*)$. A fixed point is locally stable if, given an initial condition sufficiently close to fixed point, subsequent iterates eventually approach the fixed point. This is known as locally *asymptotic* stability, as $t \rightarrow \infty$. Before the state reaches the fixed point, it may show different behavior. Behavior before the asymptotic stability is called *transient*. If the fixed point is approached for all initial conditions then the point is globally stable. For nonlinear finite-difference equations there can be more than one fixed points and when multiple fixed points are present, none of the fixed points can be globally stable. The set of initial conditions that eventually leads to a fixed point is called the *basin of attraction* of that fixed point. Often, the basin of attraction for a fixed point can have very complicated geometry. One-dimensional ordinary differential equation does not show chaos or even periodic cycles. In such an equation, the rate of change of a variable depends on its present value. In some cases the rate of change also depends to the variable at some time in past. Mackey and Glass [8] assume, a variable, e.g., x , that decays at a rate proportional to its own concentration, exponential decay, but is produced at a rate that is a function of its value some time units, e.g., ℓ , in the past. The dynamics of x then gives a *delay differential* equation,

$$\frac{dx}{dt} = P(x(t - \ell)) - \alpha x(t) \quad (2)$$

Here, function P controls x , and α is a decay constant. If P is a single-hump function in Eq. 2, the dynamics generate fixed points and chaos. Eq. 2 can then be re-written as [8],

$$\frac{dx}{dt} = \frac{0.2x(t - 20)}{1 + x(t - 20)^{10}} - 0.1x(t) \quad (3)$$

Every instantaneous state x_t of a system can be represented by a set $(x_1, \dots, x_p)$, which defines a point in the phase space [9]. A p -dimensional phase space is formed by a set of p independent linear vectors. The sequence of such points defines a curve in the phase space, called a trajectory. As time increases, the trajectories of all the state instances either penetrate the entire phase space or they converge to a lower-dimensional subset, called an attractor. The geometrical topological dimension of an attractor can be calculated from the co-relation integral [10] of the chaotic series. The dimension of the Logistic map attractor is $\simeq 0.538$, and of Mackey-Glass attractor is $\simeq 5.7$ [10]. Eq. 1, exhibits l - d chaos, and Eq. 3, exhibits h - d chaos.

3 Feature Binding Simulation

Segmentation and tracking modules [11] generate a list of video objects B_l , where $l \in L$ video objects. d_{ijl} is i^{th} *MPEG-7* descriptor of object B_l with j^{th} feature coefficient ($j \in J_i, i \in I$). J_i is number of feature coefficients in descriptor i , I is number of descriptors per object B_l , and $d_{ijl} \in \mathbb{R}$. YY_l is the feature vector for object B_l . After normalization and neighborhood interaction, we excite the

object matrix $YY_{IL} \equiv YY$ with chaotic series $f(\cdot)$. Discrete form of either a l - d chaotic series (Eq. 1) or a h - d chaotic series (Eq. 3) is used. Then, histogram based clustering, $S(\cdot)$, is applied on re-constructed trajectories of attractors [1], to generate C - $MP7$ for video objects.

For simulation, we select 5500 diverse video objects as ground truth from about 18000 frames of 30 assorted indoor and outdoor surveillance scenes. Scenes are taken from public surveillance video databases, CAVIAR, EPFL, INRS, UROCHESTER, QCIF, and also from a local, AXIS 213 PTZ, camera feed. Video objects are randomly grouped into different datasets and pre-labeled in four different classes: *has_person* (p), *has_group_of_persons* (g), *has_vehicle* (v), and *has_unknown* (u). Abrupt changes in lighting, shaking trees, shadows, and occlusions in any scene may influence the segmentation and tracking of video objects in [11]. To minimize the influence of segmentation, video objects with good and poor segmentation are both mixed together in a class. Sample video objects in different classes are available at <http://users.encs.concordia.ca/~vidprog/CMP7/Dataset>.

We select set of *MPEG-7* descriptors which offer fast extraction, interoperability, are of compact size, and are scale and resolution invariant. The selected descriptors are color layout, edge histogram, contour shape and region shape [12]. Motion can be a useful feature to distinguish moving objects in video [13], e.g., posture, variation in motion direction. However, by definition [12], *MPEG-7* motion descriptors are not suited to exhibit discrimination in pose or in postures, among target object classes, e.g., between *person* and *group_of_person*, in consecutive frames. 1) to maintain *MPEG-7* compliance for video object description, and 2) due to the absence of suitable *MPEG-7* motion descriptors, no *MPEG-7* motion descriptor is included in the target feature vector. Motion estimation for moving image regions is integrated in tracking [11].

4 Analysis

In Fig. 1, we use multidimensional scaling to visualize relative feature vector distances for different video objects in different classes for training dataset. Fig. 1 shows, compared to *MPEG-7*, *C-MP7* with h - d chaos, provides better clustering of video objects, specially for *vehicle* objects.

4.1 Multi-class Discriminant Evaluation

Multi-class Fisher criteria[14] maximizes the distance for inter-class scatter over that for intra-class scatter in different class of video objects. Considering the overall distribution of multiple classes, the weighted mean of two-class Fisher criteria contributes on the selection of the best feature for multi-classes. The μ -class ($\mu > 2$) Fisher criterion can be decomposed into w two-class Fisher criteria (where $w = {}^{\mu}C_2$), here C indicates the combination of σ -classes as pairs. The

Fig. 1. 2-D visualization of video objects in dataset 16, a) *MPEG-7* and b) *C-MP7* (*h-d* chaos)

inter-class scatter distance f_{AB} between, A with E video objects $\mathcal{Y}_A$ and class B with F video objects $\mathcal{Y}_B$, is,

$$f_{AB} = \frac{|\frac{1}{E} \sum_{A=1}^E \mathcal{Y}_A - \frac{1}{F} \sum_{B=1}^F \mathcal{Y}_B|}{\sqrt{\frac{\sum_{A=1}^E (\mathcal{Y}_A - \frac{1}{E} \sum_{A=1}^E \mathcal{Y}_A)^2}{E-1} + \frac{\sum_{B=1}^F (\mathcal{Y}_B - \frac{1}{F} \sum_{B=1}^F \mathcal{Y}_B)^2}{F-1}}} \quad (4)$$

In a μ -dimensional space, distance measure is defined as [14],

$$DIS = \min(\beta_u + \lambda \frac{1}{w} \sum_{u=1}^w (\beta_u)) \quad (5)$$

where β_u denotes the u^{th} binary class (e.g., f_{AB}) Fisher criterion among μ classes which is decomposed into w , and λ is an empirical factor. Higher distance DIS in Eq. 5 implies more discriminant feature vector for the corresponding multi-classes. Fig. 2 shows that the minimum intra-class distance, in the four classes we consider, is higher for *C-MP7* (with *h-d* chaos) than that for *MPEG-7* and for *C-MP7* (with *l-d* chaos).

4.2 Classification Accuracy

Initially [1] reports, better cross-validation accuracy can be achieved using *C-MP7* over *MPEG-7*, with either *h-d* or *l-d* chaotic series. In surveillance scenes, usually changes in video objects in successive frames are not significant. Thus, in cross-validation, the training and testing datasets may share almost similar image regions in successive frames. Here, we present classification accuracy with different training and test datasets. One pruned (no repeated object to avoid over fitting in training) dataset with few video objects (i.e., 167) is selected for training. In test datasets, repeated objects from subsequent frames are kept.

Fig. 3 shows that, significant improvement can be achieved by *C-MP7* (using *h-d* chaos). In SVM, the accuracy improves 10% to 20% over *MPEG-7* for all four classes, i.e., *p*, *g*, *v*, *u*. *C-MP7*, however, performs poor, if *l-d* chaos is used. In addition to over all classification performance in all classes, above 99% accuracy is achieved for only *vehicles* against other objects in binary SVM classifier, for *C-MP7* with *h-d* chaos. The corresponding accuracy for *MPEG-7* is on average 84%. The said high *vehicle* accuracy complements the well clustered feature vector representation of *vehicles* in Fig. 1-b.

Fig. 2. Multi-class Fisher criteria for four classes

4.3 Why High-Dimensional Chaos Is Better?

A chaotic series must be sensitive to initial conditions[8]. One reason for excellence of *C-MP7* with *h-d* chaos compared to that with *l-d* chaos, can be due to the sensitivities of initial seed x_0 , either, in the chaotic series, or in the chaotic attractor. In Fig. 4-a, up to $n=15$, *l-d* chaos (Eq. 1) exhibits same series with initial seeds of 10^{-6} difference, i.e., 0.123423 and 0.123425. As n increases, chaotic series drift apart. In 2-dimensional phase space, $x(n)$ vs $x(n+1)$, both the series, however, stay on the same attractor, see Fig. 4-b. In case of *h-d* chaos (Eq. 3), at higher iteration, e.g., $n=1980$, the said two seeds show drift in Fig. 4-c, but stay on the same attractor in Fig. 4-d. However, for initial seeds of 10^{-1} difference, *h-d* chaotic attractors show significant drift in Fig. 4-d. The non-linear structure (i.e., dynamics in trajectory) in *l-d* chaotic series is easily detectable and is hardly detectable in *h-d* chaotic series. To verify, the effect of drifts in chaotic series, we re-simulate *C-MP7* with *l-d* chaotic series at 10 iteration, where there is no drift in the series, and with *h-d* chaotic series at 2500 iteration, where there exists significant drifts in the series. As in Fig. 3, for SVM, the *C-MP7* with *l-d*

Fig. 3. Accuracy with *C-MP7* and *MPEG-7*

Fig. 4. a) *l-d* series, b) *l-d* attractors, c) *h-d* series, and d) *h-d* attractors

chaos again shows poor accuracy, while, with *h-d* chaos achieves increased accuracy, on average 84.58%. The drifts in *h-d* chaotic attractors, not in chaotic series, for change in initial seeds in the range of 10^{-1} , is due to transient. In Fig. 4-d, this transient is the initial delay before the *h-d* chaotic series rides on the attractor, and is caused by asymptotic stability of fixed points in the

Fig. 5. Variance in *MPEG-7* and *C-MP7*

attractor [8]. Transient is practically insignificant in *l-d* chaos. In a class, this transient allows *C-MP7* to capture subtle variations of descriptor coefficients, as they exist in diverse image-regions for video objects in successive frames in surveillance scenes. The higher variance in *C-MP7* with *h-d* chaos as shown in Fig. 5 for video objects, manifests the said subtle variations.

5 Conclusion

The proposed histogram-based chaotic feature binding of *MPEG-7* visual descriptors generates a new feature vector *C-MP7*, and is *MPEG-7 compliant*. *C-MP7*, with *h-d* chaotic series, video object classification accuracy is better, than the *C-MP7* with *l-d* chaos, and with original *MPEG-7*. Our analysis, validates the excellence of *h-d* chaos-based *C-MP7* as a feature vector over *MPEG-7*, for video objects, despite poor segmentation and tracking. Video objects of different scale, resolution, orientation and occlusion are randomly mixed, and tested from both indoor and outdoor surveillance scenes. *Vehicle* objects are clustered well with *C-MP7*, which leads to above 99% accuracy for only *vehicles* against other objects in all data sets.

The new feature vector *C-MP7* has the same *MPEG-7* format as available in the XML schema for each descriptor, except null feature coefficients added at varying indexes. So each descriptor for a video object can be reconstructed back to the original XML description with reduced feature coefficients. This compliance allows the *C-MP7*-based video object descriptions to be shared with other potential *MPEG-7* industry applications.

References

1. Azhar, H., Amer, A.: Chaos and MPEG-7 Based Feature Vector for Video Object Classification. In: IEEE International Conference on Image Processing, pp. 432–437 (2008)
2. Ma, X., Eric, W., Grimson, L.: Edge-based rich representation for vehicle classification. In: IEEE International Conference on Computer Vision, pp. 1185–1192 (2005)

3. Qiming, L., Khoshgoftaar, T.M., Folleco, A.: Classification of Ships in Surveillance Video. In: IEEE International Conference on Information Reuse and Integration, September 2006, pp. 432–437 (2006)
4. Eidenberger, H.: How Good are the Visual MPEG-7 Features? In: SPIE Visual Communications and Image Processing, vol. 5150, pp. 476–488 (2003)
5. Basar, E. (ed.): Chaos in Brain Function. Springer, Heidelberg (1990)
6. Hubel, D.H.: Eye, Brain, and Vision, 2nd edn. W. H. Freeman, New York (1995)
7. Skarda, C.A., Freeman, W.J.: How Brains Make Chaos in Order to Make Sense of the World. *Behav. Brain Sci.* 10, 161–195 (1987)
8. Kaplan, D., Glass, L.: Understanding Nonlinear Dynamics. Springer, Heidelberg (1998)
9. Takens, F.: Dynamical Systems and Turbulence. *Lecture Notes in Mathematics*, p. 898. Springer, Heidelberg (1981)
10. Grassberger, P., Procaccia, I.: Characterization of Strange Attractors. *Phys. Rev. Lett.* 50, 346–349 (1983)
11. Amer, A.: Voting-based Simultaneous Tracking of Multiple Video Objects. *IEEE Transactions on Circuits and Systems for Video Technology* 15, 1448–1462 (2005)
12. Manjunath, B.S., Salembier, P., Sikora, T.: Introduction to MPEG-7. John Wiley and Sons, Ltd., Chichester (2002)
13. Javed, O., Shah, M., Shafique, K.: Automated Visual Surveillance in Realistic Scenarios. *IEEE Multimedia Magazine* (2007)
14. Jack, L., Guo, H., Nandi, A.: Feature Generation Using Genetic Programming with Application to Fault Classification. *IEEE Transactions on Systems, Man and Cybernetics Part B: Cybernetics* 35(1) (2005)