

RECURRENCE TEXTURES FOR HUMAN ACTIVITY RECOGNITION FROM COMPRESSIVE CAMERAS

Kuldeep Kulkarni, Pavan Turaga

Schools of Arts, Media, Engineering, and Electrical, Computer, and Energy Engineering
Arizona State University, Tempe, AZ
Email: {kkulkar1,pturaga}@asu.edu

ABSTRACT

Recent advances in camera architectures and associated mathematical representations now enable compressive acquisition of images and videos at low data-rates. In such a setting, we consider the problem of human activity recognition, which is an important inference problem in many security and surveillance applications. We propose a framework for understanding human activities as a non-linear dynamical system, and propose a robust, generalizable feature that can be extracted directly from the compressed measurements without reconstructing the original video frames. The proposed feature is termed *recurrence texture* and is motivated from recurrence analysis of non-linear dynamical systems. We show that it is possible to obtain discriminative features directly from the compressed stream and show its utility in recognition of activities at very low data rates.¹

Index terms : Activity Analysis, Inference from Compressive Cameras

1. INTRODUCTION

Recent years have seen significant progress in the fields of compressive sensing, which allows signal reconstruction at sub-Nyquist sampling rates by exploiting additional structure on the signal being sensed. This is most often in the form of sparsity in an appropriately chosen basis [1]. A significant body of work now exists that deals with algorithms for recovery of the original signal from such compressed measurements. There is a tremendous breadth of such techniques, and we refer to recent compilations for a comprehensive survey [2, 3]. However, much less attention has been devoted to the question of whether higher-level inference tasks such as detection and recognition can be performed without reconstructing the original signal/images. Recent work shows that simpler tasks like background subtraction[4] are possible using compressive sensing without reconstruction. The general problem of activity recognition is difficult to address since

many features that are useful for object and activity recognition tasks require non-linear feature extraction techniques.[5] explored the utility of CS as a compression tool for features that have already been extracted from the original video, but did not address direct feature extraction from CS measurements of images. Typical features useful for activity analysis include histogram of gradients (HOG) [6], optical flow [7], 3D SIFT [8], contours [9] etc. Activity recognition has a rich and long history in computer vision, and we refer to recent surveys on this topic [10]. It is quite difficult to obtain such complex features directly from the compressive measurements without an intermediate step of signal reconstruction. Thus, there is a growing need to explore novel features that retain robustness and accuracy, yet are amenable to extraction directly from compressed measurements. Recently a linear dynamical system (LDS) was used to recover videos from CS cameras in [11]. LDS models are useful for video reconstruction, but being generative models they are sensitive to spatial/view transforms, thus require further processing to obtain robust recognition performance.

Since we do not wish to reconstruct images, it is interesting to consider what class of features could be preserved when the video signal is projected from the high dimensional pixel-space into a low dimensional space. In this context, we revisit the Johnson–Lindenstrauss (JL) lemma [12] which states that a small set of points in a high-dimensional space can be embedded into a much lower dimensional space in a manner that *nearly* preserves relative distances between the points. This suggests, that one might consider extracting features which encode the geometric properties of the video signal, since these properties would be preserved in the compressed domain as well. However, the question remains whether such geometric properties contain sufficient discriminative information for the purposes of activity classification.

The raw geometric relationship between points in high-dimensions is often encoded in terms of distance matrices or affinity matrices [13, 14]. However, it is in general not trivial to compare the geometries of two high-dimensional point clouds. On a related note, recurrence quantification analysis (RQA) has been recently proposed as a tool to quan-

¹This work was sponsored in part by ASU startup grant and ONR Grant N00014-12-1-0124 subaward Z868302.

tify fine variations in non-linear dynamical system parameters [15, 16, 17]. These techniques have also been widely adopted in various studies in behavioral sciences [18]. Recurrence plots are similar to affinity matrices, and are closely related to the geometric properties of the dynamical system. A similar approach has been proposed for view-robust activity recognition using temporal self-similarities [19].

Contributions: In this paper, we make the following contributions. 1) We study the problem of human activity recognition from compressive cameras using the geometric properties of high-dimensional video data, 2) We present a conceptually simple yet robust method for quantifying this geometric information in terms of recurrence textures, 3) We show the utility of this method for performing robust activity recognition at very low data rates.

Organization: The rest of the paper is organized as follows. In section 2 we provide a theoretical framework to consider the problem of human activity analysis in compressive cameras. In section 3, we discuss the proposed geometric analysis of video via recurrence analysis, and associated feature extraction. In section 4, we present experimental results, and conclusions in section 5.

2. PROBLEM FORMULATION

When a sequence of images is acquired by a compressive camera, the measurements are generated by a sensing strategy which maps the image space $\mathcal{I} \in \mathbb{R}^N$ to an observation space $Z \in \mathbb{R}^M$. The overall mapping consists of a transformation F from the 3D scene-space \mathcal{S} to image-space, with the addition of noise n in the sensor, followed by the measurement matrix ϕ , which gives measurements Z ,

$$I(t) = F \circ S(t) + n(t) \quad (1)$$

$$Z(t) = \phi I(t) \quad (2)$$

Here $S(t)$ refers to a model of the scene (such as a CAD model) with a human performing an action. Compressive sensing represents a succession of data-reduction operations, going from the full-blown space of 3D scenes to image-space, and then to measurement-space. Assuming that the changes in the scene are due to a human performing some activity, we seek features that can be extracted directly from the sequence of measurements $\{Z(t)\}$. Since we do not intend to reconstruct the image-sequence, we are restricted in our ability to extract meaningful features. However, the JL-lemma suggests that the general geometric relations of a set of points in a high-dimensional space can be preserved by certain embeddings into a low-dimensional space. In the case of compressive sensing, this embedding is achieved by the random measurement matrix ϕ , in other words orthogonally projecting to \mathbb{R}^M . The preserving of relative distances between images under such an embedding motivates us to explore the notion of recurrence plots [15]. Further, considering that the system de-

finied in (2) is a non-linear dynamical system, we consider understanding the system properties via its recurrence properties [16, 17].

3. RECURRENCE TEXTURES AND CLASSIFICATION OF ACTIVITIES

Recurrence plots (RPs) are a visualization tool for dynamical systems. A recurrence matrix defined as

$$R(i, j) = \theta(\epsilon - \|x_i - x_j\|_2) \quad (3)$$

where x_t is the observed time series and $\theta(\cdot)$ is the Heaviside step function. RPs, which are thus binary images displaying black dots where the values are within the threshold ϵ , are shown to capture the system’s behavior and be distinctive for different dynamical systems. At the time instant t , the compressive measurement of the image observation (the t^{th} frame of the video sequence) is $Z(t) \in \mathbb{R}^M$. Thus, if a sufficient number of measurements are taken, then with high probability the RPs for the compressed $\{Z(t)\}$ and uncompressed signals $\{I(t)\}$ will be the same. This is a straightforward consequence of the JL-lemma.

Thus, we propose to use the recurrence relations of $\{Z(t)\}$ as a means to acquire discriminative features from activities. In order to quantify the structures in RPs, a set of measures known as Recurrence Quantitative Analysis have been proposed by [15, 16, 17]. However, the lumped nature of RQA measures do not capture the dynamics of different system unambiguously, sometimes yielding similar RQA measures for structurally dissimilar RPs [20]. Moreover, the RPs themselves are very sensitive to the threshold, leading to different structures for different thresholds for the same system. These limitations motivate us to make use of the full geometric information encoded in the non-thresholded recurrence matrices.

We term the non-thresholded recurrence matrices simply as ‘Distance’ matrices. But instead of calculating the distance matrix for the time series obtained from the sequence of measurements, we calculate it for the time series obtained by taking the first derivative measurements (successive difference operation). Thus, for each sequence of compressive measurements $\{Z(t)\}$ the distance matrix is a square-symmetric matrix, D of size $(T - 1) \times (T - 1)$, given by

$$D(i, j) = \|\dot{Z}(i) - \dot{Z}(j)\|_2 \quad (4)$$

where $\dot{Z}(i) = Z(i + 1) - Z(i)$. We perform this successive difference operation as a way to remove the effects of a static background, so that that features are more sensitive to movement in the scene. On visualizing the distance matrices as intensity images as shown in figure 1, it is clear that different activities give rise to widely different *recurrence textures*.

Motivated by this, we pose the problem of classification of the dynamical system as a texture recognition problem. To

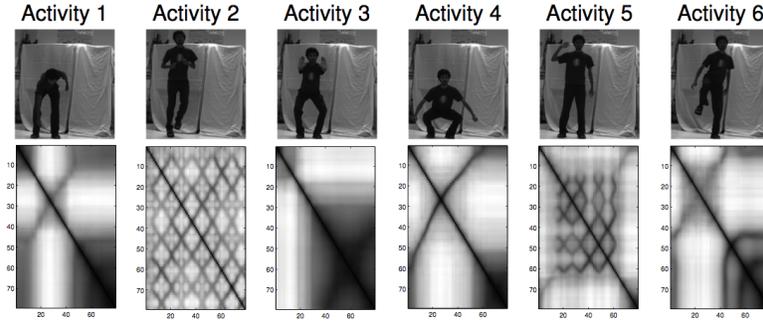


Fig. 1. Row1: Examples of different activities from UMD dataset; Row2: Corresponding recurrence texture representations of the actions.

this end, we utilize a computationally simple yet powerful texture classification method based on local binary patterns (LBPs) [21]. Certain LBPs termed as ‘uniform’ are fundamental properties of image texture and their occurrence histogram is proven to be a powerful texture feature.

4. EXPERIMENTS

For experiments, we choose the UMD Human Activity Dataset [22]. This database consists of 10 different activities: Bend, Jog, Push, Squat, Wave, Kick, Batting, Throw, Turn Sideways and Pick Phone. Each activity was repeated 10 times, so there were a total of 100 sequences in the dataset. Each sequence consists of 80 images and were cropped to a resolution of 331×301 . Each image is sensed compressively at measurement factors of 100, 400, 800, 1000 and 1200 by taking the corresponding number of random measurements. Since the background is relatively static, in each sequence, differences of compressive measurements of successive images are taken to remove the effect of the static background. These difference measurements are used to generate a distance matrix of size 79×79 for each sequence. As explained in section 3, these distance matrices are viewed as textures.

We used local binary pattern features [21] to classify the textures. Thus, each sequence is represented by LBP feature descriptor of length 38 which gives the normalized histograms of 38 binary patterns. For this experiment, we performed a leave-one-execution-out test, in which we trained on 9 executions and tested on the remaining execution for all activities using a simple nearest-neighbor classifier. In table 1, we show the confusion matrix obtained for the activity recognition experiment using the proposed method for a compression factor of 100. The classification accuracy is obtained to be 90%.

In table 2, we present average recognition results when the compression ratio was varied across a broad range of values. We observe that the proposed framework works very well across a wide variety of compression factors. These are encouraging and positive results, which suggest that significant

Activity	1	2	3	4	5	6	7	8	9	10
1	10	0	0	0	0	0	0	0	0	0
2	0	10	0	0	0	0	0	0	0	0
3	0	0	9	1	0	0	0	0	0	0
4	0	0	0	10	0	0	0	0	0	0
5	0	0	0	0	10	0	0	0	0	0
6	3	0	0	0	0	6	0	1	0	0
7	0	0	0	0	0	0	10	0	0	0
8	1	0	0	0	0	0	0	7	1	1
9	0	0	0	0	0	0	0	0	10	0
10	0	0	0	0	0	0	0	2	0	8

Table 1. Confusion table for activity recognition experiment using compressive measurements at a compression ratio = 100. The confusion matrix exhibits a strong diagonal structure, which implies that most activities are recognized correctly.

Compression factor	Recognition Rate
Uncompressed	90%
100	90%
400	86%
800	84%
1000	81%
1200	80%

Table 2. Activity recognition rate for different compression factors. The recognition rates are quite stable even at very high compression rates.

performance improvements are possible by a careful choice of features and classifiers. **The UCSD Traffic Dataset**[23] consists of 254 videos capturing traffic of three types: light, moderate, and heavy. Each video is of length 50 frames at a resolution of 64×64 pixels. We perform a classification experiment of the videos into these three categories. There are four different train-test scenarios provided with the dataset. For comparison, firstly at fixed compression ratio of $25\times$, we per-

form the same experiments with CS-LDS[24] as well as our method. The results show that our method performs signifi-

	Expt.1	Expt.2	Expt.3	Expt.4
Our method	92.06	92.19	85.94	92.06
CS-LDS(d=10)	84.12	87.5	89.06	85.71

Table 3. Classification results (in %) on the UCSD Traffic Dataset.

cantly better than CS-LDS method for the compression ratio. Secondly, we perform the 4 experiments using our method for different compression ratios.

Compression ratio	Expt.1	Expt.2	Expt.3	Expt.4
25×	92.06	92.19	85.94	92.06
150×	88.89	78.13	78.13	82.54
300×	87.30	82.81	76.56	82.54

Table 4. Classification results at different compression ratios (in %) on the UCSD Traffic Dataset.

5. CONCLUSIONS AND DISCUSSIONS

In this paper, we presented a framework to address human activity recognition from compressive cameras. This has potential applications in a wide variety of resource constrained contexts such as in remote air-borne surveillance, or home-based security and health-care systems. We proposed a solution based on dynamical analysis via recurrence relations, which has an interpretation in terms of geometric structures of high-dimensional data. We showed that these geometric structures are preserved even in the compressed domain, and do contain significant discriminative information to recognize activities at very low data-rates. This opens up several lines of further research. One question would be to consider theoretical guarantees that relate preservation of geometric structures to the proposed features. Further, we expect significant performance improvements on using more sophisticated classifiers and feature selection techniques.

6. REFERENCES

- [1] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [2] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 83–91, 2008.
- [3] Michael Elad, *Sparse and Redundant Representations - From Theory to Applications in Signal and Image Processing*, Springer, 2010.
- [4] V. Cevher, A. C. Sankaranarayanan, M. F. Duarte, D. Reddy, R. G. Baraniuk, and R. Chellappa, "Compressive sensing for background subtraction," in *Euro. Conf. Comp. Vision*, Oct. 2008.
- [5] O Concha, R. Xu, and M Piccardi, "Compressive sensing of time series for human action recognition," in *Proceedings of the 2010 International Conference on Digital Image Computing: Techniques and Applications*, DICTA '10.
- [6] Navneet Dalal and Bill Triggs, "Histograms of oriented gradients for human detection," in *IEEE Conf. Comp. Vision and Pattern Recog*, 2005, pp. 886–893.
- [7] R. Chaudhry, A. Ravichandran, G. D. Hager, and R. Vidal, "Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions," in *IEEE Conf. Comp. Vision and Pattern Recog*, 2009, pp. 1932–1939.
- [8] I. Laptev and T. Lindeberg, "Space-time interest points," *IEEE Intl. Conf. Comp. Vision.*, 2003.
- [9] A. Veeraraghavan, A. Roy-Chowdhury, and R. Chellappa, "Matching shape sequences in video with an application to human movement analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 12, pp. 1896–1909, 2005.
- [10] J.K. Aggarwal and M.S. Ryoo, "Human activity analysis: A review," *ACM Comput. Surv.*, vol. 43, pp. 16:1–16:43, April 2011.
- [11] A. C. Sankaranarayanan, P. Turaga, R. Baraniuk, and R. Chellappa, "Compressive acquisition of dynamic scenes," in *Euro. Conf. Comp. Vision*, Sep. 2010.
- [12] W. B. Johnson and J. Lindenstrauss, "Extensions of lipschitz mapping into Hilbert space," *Contemporary Mathematics*, vol. 26, pp. 189–206, 1984.
- [13] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [14] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," *Adv. Neural Inf. Proc. Sys.*, pp. 585–591, 2001.
- [15] J. P. Eckmann, S. O. Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," *Europhysics Letters*, vol. 5, no. 9, pp. 973–977, 1987.
- [16] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," *Physics Reports*, vol. 438, no. 5-6, pp. 237, 2007.
- [17] Webber Jr. C.L. Zbilut, J.P., "Embeddings and delays as derived from quantification of recurrence plots," *Physics Letters A*, vol. 171, no. 3-4, pp. 199–203, 1992.
- [18] "Tutorials in contemporary non-linear methods for the behavioral sciences," in *National Science Foundation, Webbook*, M. A. Riley and G. C. Van Orden, Eds., 2005.
- [19] Imran N. Junejo, E. Dexter, I. Laptev, and P. Pérez, "View-independent action recognition from temporal self-similarities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 1, pp. 172–185, 2011.
- [20] J. S. Iwanski and E. Bradley, "Recurrence plots of experimental data: To embed or not to embed?," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 8, no. 4, pp. 861–871, 1998.
- [21] Timo Ojala, Matti Pietikäinen, and Topi Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, 2002.
- [22] A. Veeraraghavan, R. Chellappa, and A. K. Roy-Chowdhury, "The function space of an activity," *IEEE Conf. Comp. Vision and Pattern Recog*, pp. 959–968, 2006.
- [23] A. B. Chan and N. Vasconcelos, "Probabilistic kernels for the classification of auto-regressive visual processes," in *IEEE Conf. Comp. Vision and Pattern Recog*, June 2005.
- [24] A. C. Sankaranarayanan, P. Turaga, R. Baraniuk, and R. Chellappa, "Compressive acquisition of dynamic scenes," in *under review at SIAM J. Imaging Sciences*.