IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 9, 2019, accepted October 24, 2019, date of publication October 29, 2019,
date of current version December 13, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2950313

Activity Interaction Detection by Using Causal Discovery With Order Estimation

YAWEN FAN[1], QUAN ZHOU[1], BIN KANG[2], AND XIAODONG BAI[1]

[1]Key Laboratory of Ministry of Education for Broad Band Wireless Communication and Sensor Network Technology, Nanjing University of Posts and Telecommunications, Nanjing 210003, China
[2]College of Internet of Things, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

Corresponding author: Yawen Fan (ywfan@njupt.edu.cn)

This work was supported in part by the Natural Science Foundation of Jiangsu Province of China under Grant BK20160908, Grant BK20181393, and Grant BK20170915, in part by the NUPTSF under Grant NY214139, in part by the NSFC under Grant 61876093, Grant 61881240048, Grant 61671253, Grant 61701260, and Grant 61801242, in part by the Natural Science Foundation for Jiangsu Higher Education Institutions under Grant 16KJB510031, and in part by the Jiangsu Overseas Visiting Scholar Program for the University Prominent Young and Middle-Aged Teachers and Presidents.

ABSTRACT Interaction detection is a fundamental task in video activity analysis. Activities contain group structure and temporal order information, which makes interactions complex. In this paper, a novel framework is proposed to explore the global and local activity interactions by using Granger causality discovery in multivariate time series. Based on the inherent properties of time series dependent structures related to variable time orders, an order selection algorithm considering group information is proposed. Experiments on the real world video surveillance dataset show that the activity network constructed by the proposed method is hierarchical, including global and local dependence structure.

INDEX TERMS Activity analysis, interaction detection, Granger causality, order estimation.

I. INTRODUCTION

One main problem in video data mining is to discover hidden temporal dependencies in the sequential data. In temporal data mining, the input data is typically a sequence of discrete items associated with time stamps. This study aims to automatically detect the temporal interactions between them, just using local motion features as inputs, such as optical flow vectors. For video surveillance scenes, there co-exist a number of activities. By performing the dimensionality reduction, local motion features are clustered into atom activities, which are more easily to interpret. And these atom activities can be deemed as multiple temporal variables. As a powerful tool for sequential data analysis, Granger causality [1] has played a key role in understanding behaviors in many domains to detect dependence structure of multivariate time series, including economics [2], climate science [3] and neuroscience [4]. Recently, the Granger causal framework has been successfully applied to computer vision, especially activity analysis [5]–[9].

The original concept of Granger causality was proposed by Wiener and then introduced into data analysis by Granger [1]. A variable A Granger causes another variable B, if the prediction of B is significantly improved when the historical information of A is included in the prediction model. And this notion is extended from two variables to multiple variables through the analysis of vector autoregressive (VAR) models [10].

The first challenge in using Granger causality to detect temporal dependencies is that in real-world video scenes, the interactions among activities are inherently complex. There exists structural grouping information about variables. Thus it is helpful to incorporate this information into the Granger causality framework. In a traffic scene, for example, the activities can usually be naturally grouped according to their location, direction and so on. If these structural information can be used, it is possible to achieve activity interaction analysis more accurately and efficiently

The second challenge comes to the problem of selecting time lags among activities. The temporal interaction between two activities is typically characterized within a specific time lag. For example, in cases where the time lag between related activities is assumed to be large, it can be considered that related activities with a short lag occur at the same time. However, when the time lag between related activities is viewed small, it is considered as a sequence pattern. Therefore, time

The associate editor coordinating the review of this manuscript and approving it for publication was Feng Shao.

lag is one of the key characteristics to determine the time dependence relationship.

The objective of the present article is to demonstrate that leveraging group structure among the temporal variables can indeed help improve the accuracy of Granger causality measure, and specifically the effectiveness of the proposed method is largely due to the use of regression methods with adaptive order. The main contributions of this paper are listed as follows:

1) A new framework to identify the local and global interactions is proposed. Specifically, the group structure of activities is taken into account. Both inter-group and intra-group level causalities can be measured.
2) A modified Bayesian information criterion is proposed to estimate the order of the VAR model. The time lag is adaptive to the global and local characteristics.
3) Provide a more effective method for constructing Granger causality network.

The rest of the paper is organized as follows. In Section II, we do literature review on Granger causality and order selection. In Section III, the proposed framework is introduced. In section IV, activity detection, clustering and multivariate time series generation are introduced. Section V describes the intra-group and inter-group causality analysis, and in section VI adaptive order selection algorithm is discussed in details. VII describes the experimental results in real dataset. Section VIII concludes the paper and discusses future studies.

II. RELATED WORK

Many works of temporal dependency discovery employ the framework of Granger causal models. Multivariate time series often contains structural grouping information, and this knowledge can be incorporated into the Granger causality framework through the regression algorithm with different Group Lasso penalties. Specifically, Lozano *et al.* [11] introduced a penalty called Group Boosting, to group all the past observations, in order to construct a relatively simple Granger network model; Bolsta *et al.* [12] performed variable selection using Group-lasso principles by assuming that the time series was stationary. However, these methods can only realize model selection at the group level, and cannot enable the selection of variables within the group. To overcome this issue, Basu *et al.* [13] introduced the concept of direction consistency in the Group Lasso regression regularization framework, in order to realize the variable selection at the inter-group and intra-group level, when constructing the network Granger causality for panel data. However, in these works the time lag dependence structure, which exists in multi-variant time series, has not been taken into account.

Based on the inherent property of time series dependent structures related to variable time orders, a number of works focus on order selection for modeling multivariate time series. A popular approach is to try different values for the lag and then to choose a value based on some criteria such as Bayesian information criterion (BIC). For example,

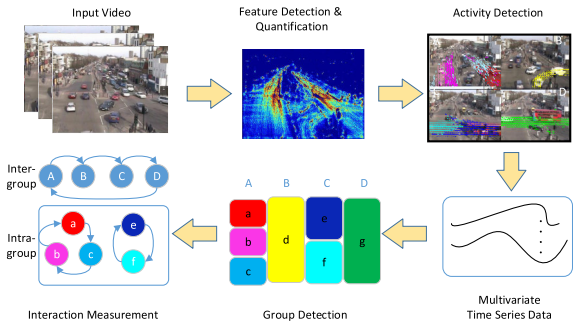


FIGURE 1. The proposed framework.

Vlachos and Kugiumtzis [14] proposed a backward-in-time-selection (BTS) method to select the lag order of the VAR model for multivariate time series. Furthermore, Siggiridou *et al.* [15] adapted the BTS method for Granger causality, in which the BIC is used as criterion. Du *et al.* [16] proposed a framework which can infer the time lag and causality at the same time. One of the main issues with this approach is that, it cannot incorporate additional prior information that the user might have. The prior information, when modeled appropriately, can aid in better estimating the maximum lag and improve the performance of causal discovery.

III. PROPOSED FRAMEWORK

A. FRAMEWORK

Considering the importance of grouping information and the parameter of time lag during time series analysis, this study proposes a framework (Fig.1) that can integrate group detection and lag estimation, which would provide more accurate causality analysis. Firstly, the low-level features are extracted and quantified. The activities are then detected based on the topic model. Each activity is further treated as a time variable to generate a multivariate time series. In order to obtain the group structure information, the activities are clustered into groups by using the K-means. Finally, through the vector autoregressive (VAR) model, Granger causality is used to measure interactions within and between groups. A new method is proposed to estimate the time order of VAR model by taking into account the group information. Based on the analysis of causality within and between groups, the hierarchical network of the whole scene can be inferred.

B. ASSUMPTION

Suppose the video scene contains H activity groups, and each group has multiple activities. The purpose of this study is to discover time dependencies between groups and between individuals in each group. As shown in Fig. 2, it is an example of a video scene which includes six activities that can be clustered into two groups. Let Y_A and Y_B represent two activity groups, and the start of Y_A causes the start of Y_B after time interval τ_G. It indicates that the occurrence of Y_B depends on the occurrence of Y_A, which means that an item within Y_A

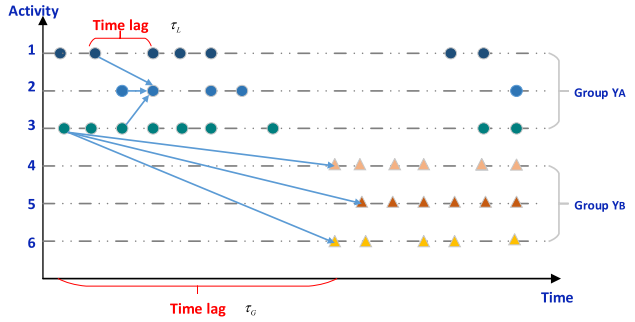


FIGURE 2. An example of activity interactions.

is often followed by an item within Y_B. At the same time, within the groups, there also exist temporal dependencies among activities with different time lags. It is obviously that interactions between and within groups are with different time scales. Interactions between groups are within large time intervals, while internal interactions are within finer intervals. Based on this prior knowledge, this study proposes an adaptive time order selection criterion for VAR models.

IV. MULTIVARIATE TIME SERIESE GENERATION

A. ACTIVITY DETECTION

In this section, the topic model HDP is adopted to model video scenes, in which latent topics are deemed as activities [17]–[19]. HDP is a nonparametric hierarchical Bayesian model that can automatically discover the number M of topics [20], as shown in Fig.3. Firstly, each video is temporally divided into N non-overlapping clips. And then optical flow vectors are extracted and quantified into visual words according to the directions and positions. Furthermore, visual documents are generated by the words accumulation over the corresponding video clips. Finally, M activities can be automatically detected by the HDP model and activity-proportion G_n for every clip form a matrix. This matrix is an M-dimensional multivariate activity-proportion series with N clips.

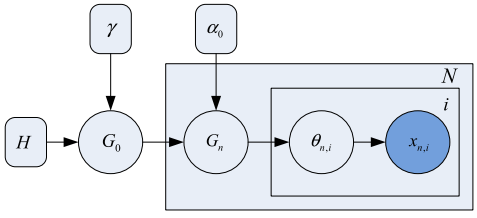


FIGURE 3. Hierarchical Dirichlet process model.

B. MULTIVARIATE TIME SERIES

By considering each activity as a time variable, each row of the activity-proportion matrix can be deemed as a discrete time series. After preprocessing, M zero-mean stationary discrete time series of length N,$\{X_1(n), ...X_M(n)\}$, $1 \le n \le N$, are generated, which can form a multivariate stationary process $X(n) = [X_1(n), ...X_M(n)]^T$. Before detecting the local and global causal relationships between activities, it is necessary to take appropriate grouping strategies to obtain the structure information. In this section, K-means clustering is used to partition these M observations into H groups, so that activities within the same group have similar dynamics, but are different from the activities of other groups. Let us consider H discrete time vector-valued stochastic processes $\{Y_1(n), ...Y_H(n)\}$ of dimensions $\{M_1, ..., M_H\}$, such that each group is a vector-valued stochastic process composed of M_h zero-mean scalar-valued stationary processes $Y_h(n) = [X_{h1}(n), ...X_{hM_h}(n)]^T$. When global interactions are being identified, each group can be treated as a whole. But when local interactions are being identified, individualities of group's members should be emphasized.

V. INTERACTION MEASURE BASED ON GRANGER CAUSALITY

A. INTER-GROUP GRANGER CAUSALITY

The intent of global analysis is to assess the interactions between groups by means of Granger Causality in the frequency domain. Given that groups are well divided, this procedure can perform without group variable selection. Different from the scalar-value based vector auto-regression, the overall process is a vector-valued process: $Y(n) = [Y_1(n), ..., Y_H(n)]^T$, and it can be represented by the VAR model,

$$Y(n) = \sum_{k=1}^{p_g} A(k)Y(n-k) + U_1(n), \tag{1}$$

where $\mathbf{A}(k)$ represents the $M \times M$ matrix of the model coefficients, $U_1(n)$ is the $M \times 1$ vector of zero means white noise, and P_g is the maximum order which needs to be decided. The spectral representation of the VAR process can be obtained by Fourier transform. The coefficients in the frequency domain is defined as,

$$\mathbf{A}(w) = \sum_{k=1}^{p_g} \mathbf{A}(k)e^{-jwkT}, \tag{2}$$

where w represents angular frequencies, and T represents the sampling period. The $M \times M$ spectral density matrix of the overall process $\mathbf{S}(w)$ and its inverse $\mathbf{P}(w) = \mathbf{S}(w)^{-1}$ are factorized as follows [21]:

$$\mathbf{S}(w) = \mathbf{T}(w)\mathbf{\Sigma}\mathbf{T}(w)^*, \tag{3}$$

$$\mathbf{P}(w) = \bar{\mathbf{A}}^*(w)\mathbf{\Sigma}^{-1}\bar{\mathbf{A}}(w), \tag{4}$$

where $\mathbf{T}(w)$ is the transfer function between variables and $\mathbf{\Sigma}$ is the noise process covariance. Because there are H groups that are considered as whole, $\mathbf{A}(w)$, $\mathbf{P}(w)$ and $\mathbf{\Sigma}$ can be broken into $H \times H$ blocks,

$$\mathbf{A}(w) = \begin{pmatrix} \mathbf{A}_{11}(w) & \dots & \mathbf{A}_{1H}(w) \\ \dots & \dots & \dots \\ \mathbf{A}_{H1}(w) & \dots & \mathbf{A}_{HH}(w) \end{pmatrix}$$

$$\mathbf{P}(w) = \begin{pmatrix} \mathbf{p}_{11}(w) & \dots & \mathbf{p}_{1H}(w) \\ \dots & \dots & \dots \\ \mathbf{P}_{H1}(w) & \dots & \mathbf{P}_{HH}(w) \end{pmatrix}, \tag{5}$$

where the dimension of the $i-j$ block is $M_i \times M_j$.

And then in order to measure the direct causality between these groups, the frequency domain causality measure for multiple vector-valued process [22] is adopted. The direct causality from Y_j to Y_i is measured at the frequency w as follows:

$$g_{j->i}(w) = \ln \frac{\left|\mathbf{P}_{jj}(w)\right|}{\left|\mathbf{P}_{jj}(w) - \bar{\mathbf{A}}^{*}_{ij}(w)\boldsymbol{\Sigma}^{-1}_{ii}\bar{\mathbf{A}}_{ij}(w)\right|}, \tag{6}$$

where $\mathbf{A}_{ij}(w)$ represents the $i-j$ block of $\mathbf{A}(w)$, $\mathbf{P}_{jj}(w)$ represents the $j-j$ block of $\mathbf{P}(w)$, and $\boldsymbol{\Sigma}_{ii}$ represents the $i-i$ block of $\boldsymbol{\Sigma}$. The total causal causality can be obtained by integrating (6) with respect to the frequency w as follows,

$$G_{inter}(j,i) = \begin{cases} \sum\limits_{w} g_{j->i}(w), & \forall\, i \neq j \\ 0, & i \neq j \end{cases} \tag{7}$$

B. INTRA-GROUP GRANGER CAUSALITY

For local analysis, the aim is to measure the direct Granger causality between each pair of activities within a group. For the group's corresponding scalar-value multivariate process: $Y_h(n) = [X_{h1}(n), ...X_{hM_h}(n)]^T$, its VAR representation is as follows:

$$Y_h(n) = \sum_{k=1}^{p_l} \mathbf{B}(k)Y_h(n-k) + U_2(n), \tag{8}$$

where $\mathbf{B}(k)$ represents the $M_h \times M_h$ matrix of the model coefficients, $U_2(n)$ is the $M_h \times 1$ vector of zero-means white noise. p_l is the order of the VAR model which will be selected according to an adaptive algorithm in the next section.

Based on the spectral representation of the VAR model, the direct Granger causality from x_{hv} to x_{hu} at frequency w is as follows:

$$g_{hu->v}(w) = \ln\left(\frac{P_{vv}(w)}{P_{vv}(w) - \left|\bar{b}_{uv}(w)\right|^2/\delta^2_{uu}}\right), \tag{9}$$

where $P_{vv}(w)$ is the $v-v$ element of $\mathbf{P}(w)$, δ^2_{uu} is the $v-v$ element of $\boldsymbol{\Sigma}$, $b_{uv}(w)$ is the $u-v$ element of $\mathbf{B}(w)$, and $1 \leq u \leq M_h$, $1 \leq v \leq M_h$. A total measure of the causality between processes X_{hv} and X_{hu} can be obtained by integrating (9) with respect to the frequency as follows:

$$G_{intar}(v,u) = \begin{cases} \sum\limits_{w} g_{hv->u}(w), & \forall\, u \neq v \\ 0, & \forall u = v \end{cases} \tag{10}$$

VI. ADAPTIVE ORDER SELCETION

The sensitivity of the results of Granger causality test to the choice of the order (time lag) is a topic of active research. Order selection can have different results for multivariate signal processing. Based on the assumption that the time lag of local interactions is always no greater than that of group interactions, an adaptive order selection method is proposed in this section.

Algorithm 1 Algorithm 1 Estimating Model Orders

Input: $Y_1, ...Y_H$: time series for each group
H : The number of groups
$P_{\max}$: The maximum order for the whole process
α_g, α_l : The penalty factor for global and local analysis
Initialize: $P_{\max}, \alpha_g, \alpha_l$
1. For p **=1:** $P_{\max}$**do // for global analysis**
2. Regress Y **for order** p **and compute** ***RSS(p)***
3. compute $mBIC(p)$ **using (12)**
4. End for
5. $[m$**BICmin**$, p_g] \leftarrow \min(mBIC)$ **using (14)**
6. $P_{L\max} \leftarrow p_g$ **// for local analysis**
7. $h = 1$
8. While $h \leq H$ **do**
9. For $p = 1$**:** $P_{L\max}$**do**
10. Regress Y_h **for order** p **and compute** ***RSS(p)***
11. Compute $mBIC(p)$ **using (12)**
12. End for
13. $[m$**BICmin**$(h), p_l(h)] \leftarrow \min(mBIC)$ **using (14)**
14. $h \leftarrow h + 1$
15. End while
Output: p_g, p_l

In practice, given time series data, a finite order p of the best VAR is selected based on theoretical principles. It is usually selected from candidate orders between 0 and $p_{\max}$. Based on the assumption in section III, $p_{\max}$ is set large enough for global analysis. As for local analysis, $p_{\max}$ is set equal to the selected best order for the global. That is,

$$p_{L\max} = p_g \tag{11}$$

Then a new criterion based on Bayesian information criterion is defined to determine the optimal value of orders,

$$mBIC(p,\alpha) = N\log(RSS(p)) + \alpha\log(N)M^2p, \tag{12}$$

where M is the number of variables, N is the length of variables, and α is a penalty factor. RSS is the regression residual sum of squares and calculated by:

$$RSS(p) = \sum_{n=1}^{N}(Y_n - \widehat{Y}_n(p))^2, \tag{13}$$

where $\widehat{Y}_n(p)$ is the prediction value calculated by regression function with different $p \in [0, p_{\max}]$. For cases in which variable length values are large and the number of variables is small, the highest candidate order is usually selected. Therefore, for local analysis, higher values of α should be specified in order to select lower orders. We compute the mBIC scores using (12) and select the optimal value of p that yields the smallest score as the optimal value,

$$p_o = \underset{0\leq p\leq p_{\max}}{\arg\min}\ mBIC(p,\alpha). \tag{14}$$

VII. ANALYSIS OF INTERACTION IN REAL DATASET

A. EXPERIMENT DATASET AND PREPROCESSING

In this section, the proposed framework was evaluated on two traffic video sequences selected from the QMUL dataset (360×288, 25 fps, 1 h). As shown in Fig. 4, these two scenes are governed by traffic lights in a certain temporal order. Typically, a scene is divided into several traffic states, each of which lasts for a period of time. For the intersection scene, there are four states, and the temporal order between them is {A->B->C->D->A...}. For the roundabout scene, there are three states, and the temporal order between them is {A->B->C->A...}.

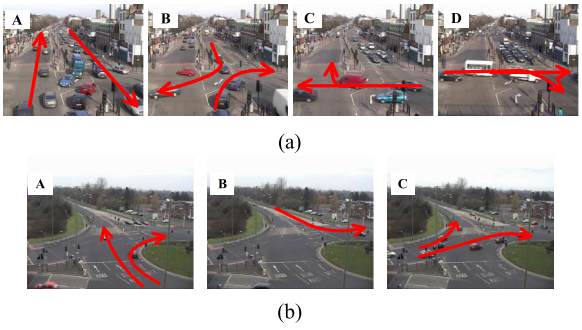


FIGURE 4. Dataset (a) intersection scene (b) roundabout scene.

To apply the topic model on video data, the video sequences are represented as bag-of-words. Firstly, the scene was spatially divided into a 36×29 grid with a spacing of 10 pixels, and then, the motion direction was quantified into 8 orientations for each cell. Therefore, a codebook with $36 \times 29 \times 8$ visual words was constructed. The whole video sequences were temporally segmented into 3-s-long non-overlapping video clips. Optical flow vectors with the magnitudes less than 0.8 were ignored and quantified into visual words. Video documents were generated by accumulating visual words. The parameters of the HDP model for the two scenes are set as follows: $\{\gamma = 0.25,\ \alpha_0 = 2,\ D_0 = 60\}$ for the intersection scene and $\{\gamma = 0.1,\ \alpha_0 = 0.5,\ D_0 = 30\}$ for the roundabout scene, in which D_0 the parameter for the Dirichlet distribution $H = Dir(D_0)$. Gibbs Sampling was used for inference in the model. HDP can automatically determine the number of topics. Finally, twenty-one and twenty-six topics (activities) were respectively discovered for the two scenes. In this study just the dominated activities, which explained at least 5% of the all flow vectors, were selected to be further analyzed and shown in Fig. 5. For the intersection scene, there were 8 activities, while for the roundabout scene, 7 activities were selected. Based on the K-means, these selected activities were clustered into groups. The number of cluster centers was manually set to be the same as the number of traffic states. For the intersection scene, activities 1, 2, 3 are clustered into the same group, and activities 6 and 7 belong to the same group. Activity 5 and activity 8 represent respectively two groups. For the roundabout scene, activities 1, 2 and 3 are in the same group,

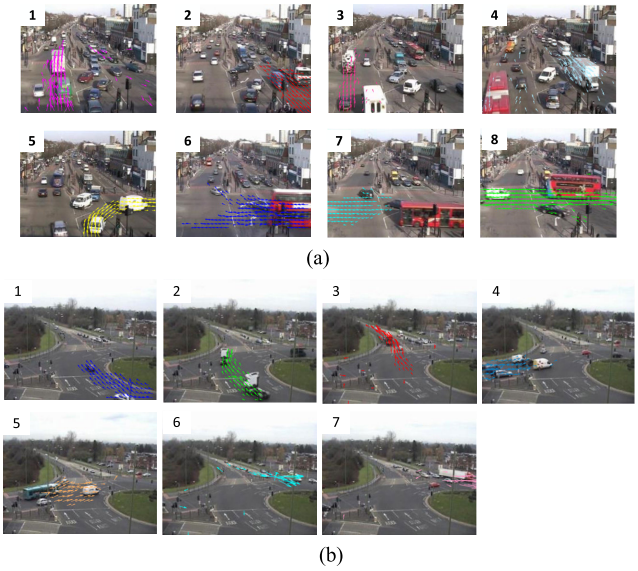


FIGURE 5. Activities detected by HDP (a) intersection scene (b) roundabout scene.

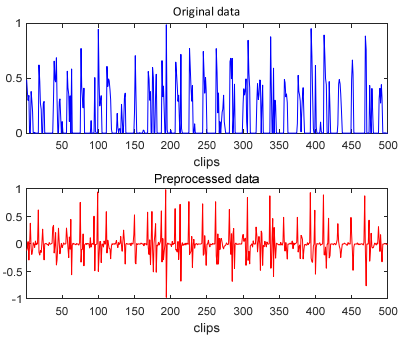


FIGURE 6. Time series of one activity.

activities 4 and 5 belong to the same group, and activities 6 and 7 belong to the same group. Obviously, the clustering results are consistent with the actual traffic statuses. Therefore, global interaction is to detect transitions between different groups. The purpose of local analysis is to discover the interactions between activities within the same group.

In light of the use of HDP model as a multivariate detection step, the distribution over activities associated with each clip formed an activity-proportion matrix. Each row represents the proportions of one activity occurring over time. In order to perform Granger causality analysis, the time series were made stationary by de-trending and differencing. As Fig. 6 shows, the top is the original time series of one activity and the bottom is the post-preprocessed data.

The coefficients and input covariance of VAR model were estimated by least squares method. Then the causal relationship between the time series is calculated from the coefficient estimates in frequency domain.

B. EFFECT OF MODEL ORDER

In this section, in order to validate the assumption that dependent structures are affected by orders of the VAR model,

the entire process is treated as a scalar-valued multivariable process and the Granger causality between pairs of activities are computed.

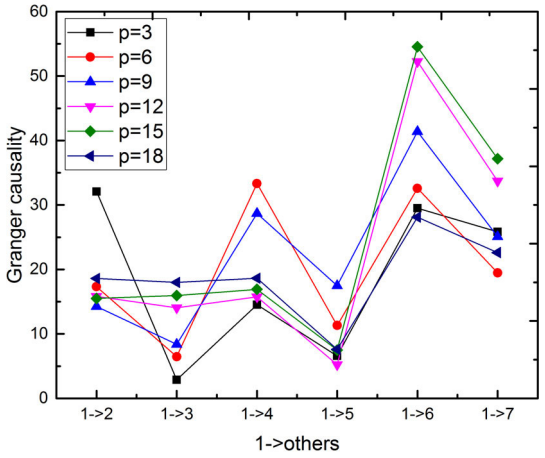


FIGURE 7. Granger causality from activity 1 to others with different orders in the roundabout scene.

Fig. 7 shows the results of causality measure from activity 1 to others in the roundabout scene. It is obviously that the value of causality is closely related with the model order. For instance, when $p = 3$, the causality from activity 1 to 2 is the strongest. But as the value of order grows, the interaction from activity 1 to activity 2 become weaker, while that from activity 1 to 7 become stronger. As we known, activity 1 and 2 belong to the same group, thus the interaction between them is local. Activity 1 and 7 belong to different groups, the time lag between them is large. This results demonstrate that it is important to adapt different scale of model orders for global and local interaction detection.

C. INTER-GROUP INTERACTION

In this subsection, the global interaction, that is, inter-group interaction, is detected. The parameters are as follows: $\{p_{\max} = 20, p_g = 7, N = 1244, H = 3\}$. For comparison, we also computed the sum of pairwise causality from group i to group j as follows:

$$G(i,j) = \sum_{l}^{M_i} \sum_{k}^{M_j} g(l,k), \tag{15}$$

where M_i represents the number of variables in group i, M_j represents the number of variables in group j, and $g(l,k)$ denotes the measure of Granger causality from activity l to k. The order is set to same as the proposed method. But in this method, the overall process is deemed as a scalar-value process. Causality between activities were measured, and summed based on the group information.

Fig. 8 shows the results of the two method for the roundabout scene. The average of the causal measurement is used as the threshold. If the result is greater than the threshold, it is considered that there is an interactive relationship between

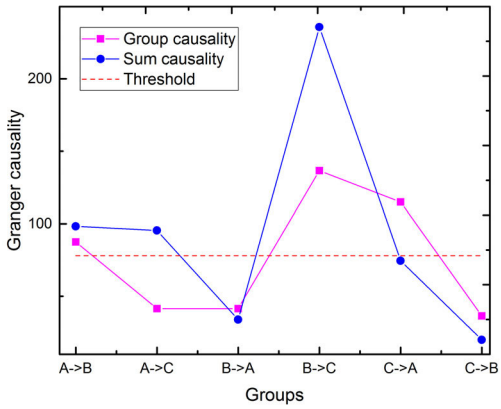


FIGURE 8. Granger causality between groups for the roundabout scene with $P_g = 7$.

the two groups. It is obvious that the complete traffic light cycle $\{A \rightarrow B \rightarrow C \rightarrow A...\}$ is discovered by the proposed methods. But the sum causality approach failed to detect the relationship between the group C and A. On the other hand, it is difficult to distinguish between direct and indirect interactions from group A to group C.

D. INTRA-GROUP INTERACTION

In this subsection, the local interactions within each group are analyzed. For the roundabout scene, seven activities are classified into three groups corresponding to the three traffic states. Group A consists of three activities, and Groups B and C include two activities respectively. Fig. 9 and Fig.10 show the results, in which the X-axis (horizontal axis) represents the frequency, while the Y-axis (vertical axis) represents the values of causal measurement. Obviously, causal measures are asymmetrical, which means that the time dependence between activities is directional. If $G_{intar}(v,u)$ is much greater than $G_{intar}(u,v)$, we believe that activity v Granger causes activity u. Specifically, there exists a local temporal order between these activities in group A:$\{1 \rightarrow 2 \rightarrow 3 \rightarrow 1\}$.

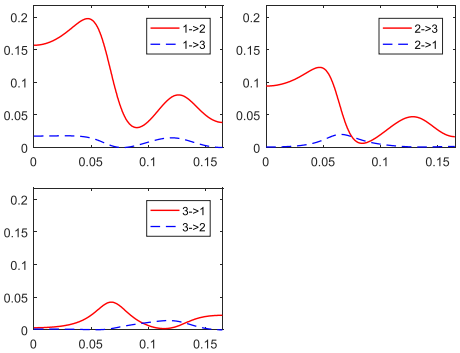


FIGURE 9. Spectral causality for each pair in group A with $p_l = 4$.

Take the second group as an example, we further demonstrate the effectiveness of the proposed order selection mehotd. Fig. 11 compares the value of mBIC and BIC varying

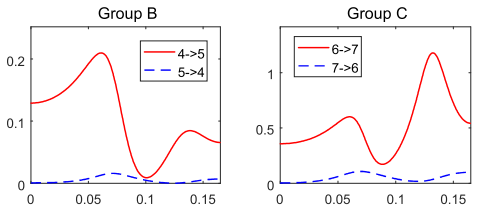


FIGURE 10. Spectral causality for each pair in group B and C with $p_l = 4$.

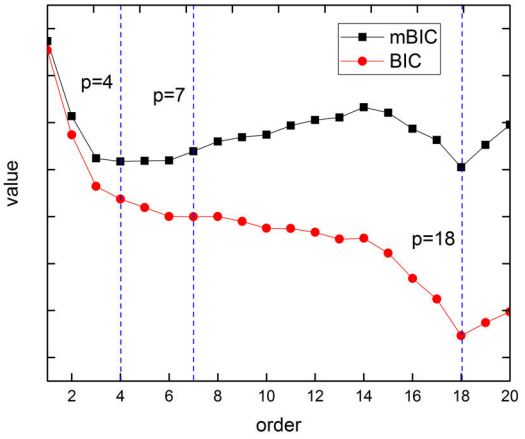


FIGURE 11. Value of mBIC and BI C varying with order.

TABLE 1. Granger causality with different orders.

Group B	*P=4*		*P=18*	
	Acivity 4	Acivity 5	Acivity 4	Acivity 5
Activity 4	0	54.5	0	35.9
Activity 5	2.7	0	35.7	0

with orders. In this test, for the proposed mBIC, we set $\alpha = 1.7$. It can be seen that there are two local minimums on the mBIC curve, while for the BIC curve there is only one. If we set $p_{\max} = 20$, then both of the two criteria will choose $p = 18$ as the optimal order. If we set $p_{\max} = 7$, which is the optimal order for the inter-group anlysis, mBIC will choose $p = 4$. But BIC will choose $p = 7$.

Table 1 shows the Granger causlity between the two activities in the group B. When $p = 18$, the causal measures in both directions are basically equal, so the two are considered to be simultaneous. When $p = 4$, the causal measure from activity 4 to activity 5 is much larger than that of the opposite direction, so the two can be regarded as sequence occurrences. As Fig. 12 shows, the experimental results are consistent with the actual situation.

E. APPLICATION: INFERING NETWORK

Table 2 and Table 3 are the causal matrixes for the two video scenes, in which gray blocks represent the local causal matrixes, and the rest global causalities. Furthermore,

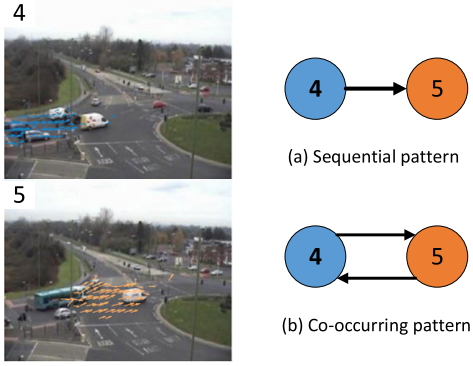


FIGURE 12. Sequential and co-occurring motion patterns.

TABLE 2. Causal matrix for the ROUNABOUT scene.

$p_g = 7$, $p_l = 4$		A			B		C	
		1	2	3	4	5	6	7
A	1	0	53.5	5.2	87.6		41.5	
	2	3.1	0	29.9				
	3	7.9	2.6	0				
B	4	41.7			0	54.5	136.6	
	5				2.7	0		
C	6	114.9			36.5		0	277.7
	7						25.7	0

TABLE 3. Causal matrix for the INTERSECTION scene.

$p_g = 8$, $p_l = 2$		A				B	C		D
		1	2	3	4	5	6	7	8
A	1	0	29.8	8.8	9.4	28.3	64.6		20.4
	2	6.9	0	10.0	1.4				
	3	11.4	0.8	0	0.8				
	4	5.9	12.5	5.1	0				
B	5	37.2				0	10.6		11.1
C	6	18.6				23.2	0	34.7	49.0
	7						18.9	0	
D	8	48.2				2.24	16.3		0

the average causal measure is used as a threshold; if the causal measure between actions/groups is greater than the threshold, there exsits an interaction between them and vice versa. The red and blue values in the tables represent causal measures greater than the threshold.

Based on the obtained causal matrixes, hierarchical networks of the whole scenes can be inferred. As shown in Fig. 13-14, it is evident that the proposed framework can identify the traffic light cycle governing the scene. And at the same time, the network of the activities within the groups are also constructed. Specifically, in intersection scene, state A is sometimes interspersed with state B.

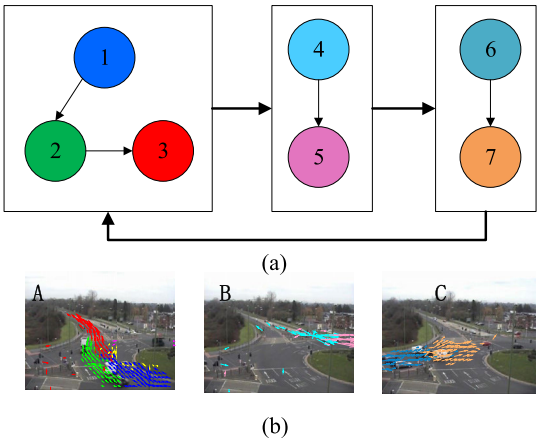


FIGURE 13. Granger causality approach applied to the roundabout scene. (a) Top: activity interactions network; (b) Bottom: visualization of the scene with three states. Different color represents different activities.

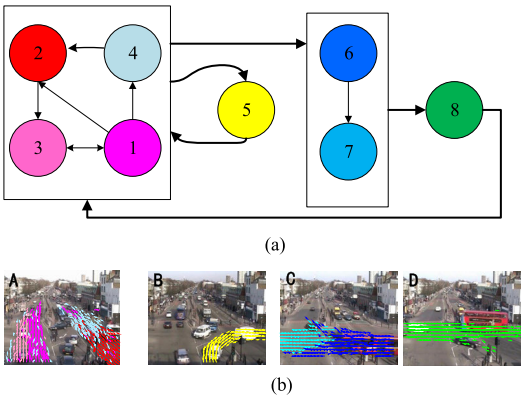


FIGURE 14. Granger causality approach applied to the intersection scene. (a) Top: activity interactions network; (b) Bottom: visualization of the scene with four states. Different color represents different activities.

VIII. CONCLUSION

In this paper, we have proposed a framework to identify the global and local interactions of activities in video scenes by detecting the Granger causal relationship based on VAR model. Activities are detected and converted into multivariate time series which can be regressed by VAR model. The framework differs from existing methods in two ways: considering both the grouping information and the parameter of time lag. For estimating the VAR model order, a new criterion called modified BIC is introduced. The experiment results have shown that the proposed adaptive order selection method can select more accurate orders. Based on the obtained global and local causal matrix, the activity network can be inferred.

In future work, the problem of overlap between groups should be taken into account. Because some activities could be shared with several groups. It would also be beneficial to study the nonlinear parametric model for Granger causality of time series.

REFERENCES

[1] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, Aug. 1969.

[2] A. Alper and O. Oguz, "The role of renewable energy consumption in economic growth: Evidence from asymmetric causality," *Renew. Sustain. Energy Rev.*, vol. 60, pp. 953–959, Jul. 2016.

[3] X. Chen, Y. Liu, H. Liu, and J. G. Carbonell, "Learning spatial-temporal varying graphs with applications to climate data analysis," in *Proc. 24th AAAI Conf. Artif. Intell.*, Atlanta, GA, USA, 2010, pp. 425–430.

[4] M. Ding, Y. Chen, and S. L. Bressler, "Granger causality: Basic theory and application to neuroscience," in *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications*. Weinheim, Germany: Wiley, Sep. 2006, pp. 437–460.

[5] K. Prabhakar, S. Oh, P. Wang, G. D. Abowd, and J. M. Rehg, "Temporal causality for the analysis of visual events," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, San Francisco, CA, USA, Jun. 2010, pp. 1967–1974.

[6] S. Yi and V. Pavlovic, "Sparse Granger causality graphs for human action classification," in *Proc. 21st Int. Conf. Pattern Recognit. (ICPR)*, Tsukuba, Japan, 2012, pp. 3374–3377.

[7] E. Swears, A. Hoogs, Q. Ji, and K. Boyer, "Complex activity recognition using Granger constrained DBN (GCDBN) in sports and surveillance video," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 788–795.

[8] D. Kular and E. Ribeiro, "Analyzing activities in videos using latent Dirichlet allocation and Granger causality," in *Proc. Int. Symp. Vis. Comput.*, vol. 9474, Dec. 2015, pp. 647–656.

[9] Y. Fan, Q. Zhou, W. Yue, and W. Zhu, "A dynamic causal topic model for mining activities from complex videos," *Multimed Tools Appl.*, vol. 77, no. 9, pp. 10669–10684, May 2018.

[10] H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*. Berlin, Germany: Springer, 2005.

[11] A. C. Lozano, N. Abe, Y. Liu, and S. Rosset, "Grouped graphical Granger modeling methods for temporal causal modeling," in *Proc. 15th ACM SIGKDD*, Paris, France, 2009, pp. 577–586.

[12] A. Bolstad, B. D. Van Veen, and R. Nowak, "Causal network inference via group sparse regularization," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2628–2641, Jun. 2011.

[13] S. Basu, A. Shojaie, and G. Michailidis, "Network Granger causality with inherent grouping structure," *J. Mach. Learn. Res.*, vol. 16, no. 1, pp. 417–453, Jan. 2015.

[14] I. Vlachos and D. Kugiumtzis, "Backward-in-time selection of the order of dynamic regression prediction model," *J. Forecasting*, vol. 32, no. 8, pp. 685–701, Jul. 2013.

[15] E. Siggiridou, V. K. Kimiskidis, and D. Kugiumtzis, "Dimension reduction of frequency-based direct Granger causality measures on short time series," *J. Neurosci. Methods*, vol. 289, pp. 64–74, Sep. 2017.

[16] S. Du, G. Song, and H. Hong, "Collective causal inference with lag estimation," *Neurocomputing*, vol. 323, pp. 299–310, Jan. 2019.

[17] X. Wang, X. Ma, and W. E. L. Grimson, "Unsupervised activity perception in crowded and complicated scenes using hierarchical Bayesian models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 31, no. 3, pp. 539–555, Mar. 2008.

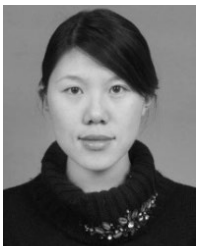
[18] D. Kuettel, M. D. Breitenstein, L. Van Gool, and V. Ferrari, "What's going on? Discovering spatio-temporal dependencies in dynamic scenes," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, San Francisco, CA, USA, Jun. 2010, pp. 1951–1958.

[19] V. Kaltsa, A. Briassouli, I. Kompatsiaris, and M. G. Strintzis, "Multiple Hierarchical Dirichlet Processes for anomaly detection in traffic," *Comput. Vis. Image Understand.*, vol. 169, pp. 28–39, Apr. 2018.

[20] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Hierarchical Dirichlet processes," *J. Amer. Statist. Assoc.*, vol. 101, no. 476, pp. 1566–1581, Dec. 2006.

[21] M. R. Gevers and B. D. O. Anderson, "Representations of jointly stationary stochastic feedback processes," *Int. J. Control*, vol. 33, no. 5, pp. 777–809, May 1981.

[22] L. Faes and G. Nollo, "Measuring frequency domain Granger causality for multiple blocks of interacting time series," *Biol. Cybern.*, vol. 107, no. 2, pp. 217–232, Apr. 2013.



YAWEN FAN received the B.E. and M.S. degrees in electronic engineering from Hohai University, Nanjing, China, in 2003 and 2006, respectively, and the Ph.D. degree from the Department of Electrical Engineering from Shanghai Jiao Tong University, Shanghai, China. She is currently an Assistant Professor with the Nanjing University of Posts and Telecommunications, Nanjing. Her research interests include computer vision, and pattern recognition and its applications in intelligent video surveillance.

QUAN ZHOU received the M.S. and Ph.D. degrees in communication and information system from the Huazhong University of Science and Technology, China, in 2006 and 2013, respectively.

He is currently an Associate Professor with the Nanjing University of Posts and Telecommunications. His research interests include computer vision and pattern recognition. He has published over 20 research articles in SCI journals, including the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON MULTIMEDIA, and *Pattern Recognition*. He has published over 20 research articles in image processing and computer vision in conference such as ICIP, ICASSP, ACCV, and ICPR. He currently serves as a TPC Member or the Chair for many international conferences. He serves as a Reviewer for a series of SCI journals, including the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON CIRCUITS SYSTEM FOR VIDEO TECHNOLOGY, *Pattern Recognition*, and *Neurocomputing*.

BIN KANG received the M.S. degree in circuits and systems from Lanzhou University, in 2011, and the Ph.D. degree in electrical engineering from the Nanjing University of Posts and Telecommunications, in 2016. He is currently a Lecturer with the College of Internet of Things, Nanjing University of Posts and Telecommunications. His research interests include computer vision and pattern recognition.

XIAODONG BAI received the Ph.D. degree in control science and engineering from the School of Automation, Huazhong University of Science and Technology, Wuhan, China, in 2014. He is currently a Lecturer with the School of Telecommunication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing, China. His research interests include machine learning, automatic monitoring of crop growth, computer vision, and embedded systems.