Graph-Based Multi-Sensor Fusion for Event Detection

Jiangfan Feng[1,2], Mingbo Gou[1] and Haibin Hu[3]

[1]School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[2]Key Lab of Instrument Science and Dynamic Test, North University of China, Taiyuan 030051, China
[3]Experiment Center, China West Normal University, Nanchong 637002, China
Correspondence should be addressed to Jiangfan Feng, fengjf@cqupt.edu.cn

Abstract

In event detection, especially where Heterogeneous Sensor Networks is used, performance of information fusion must remain high and level. In this paper, we develop a Graph-Based Multi-Sensor Fusion model for intelligent alarm in event detection. In the proposed model, we explore Graph-Based data fusion algorithms based on data-driven aspects of fusion, in order to locate the optimal sensing ranges of sensor nodes in a complex network with heterogeneous targets. In addition, we propose a framework incorporates useful contextual and temporal cues for detection alarm, exploring its conceptualizations, benefits, and challenging aspects, leading to correlation analysis of the target motion elements in the multimedia sensor stream. The aim is to give the simplest method for classical results in event detection.

Keywords: Multimedia Sensor Network, Event Detection, Directed Graph, Data Fusion

1. Introduction

Generally, event detection relies on events both in the case of impending events and of events that have occurred. Although event detection has long been a research topic, the characteristics of Heterogeneous Sensor Networks make it a non- trivial task. These disasters are a function of some unpredict reasons, such as unruly crowd aroused by some emergencies or natural hazard came from nowhere. Besides, in the modern unfortunately the most tragic disasters were allegedly due to human, entrenched invisible terrorists always find a way to break the social regulatory life. Accordingly, there is a critical need for offering early detection and alarm, and such guarantee may enable instant reaction to winning most time for evacuation and deploy, also decrease the possibility of loss.

Information fusion, since it bore in 1970s as the requirement of military, shows a considerable potential, and was designed to meet the needs of the military in the field of battle [1]. Multi-sensor data fusion (MSDF) arises at the historic moment. With the rapid development in 1990s, its technology field expanded from to military to civilian that lead to more large-scale scholarly discussion. Additionally make the disaster prevention possible. However, with a view to past experience the MSDF technology merely was restricted to combine vast correlation sensor data in some particular detection, i.e., one of the interest topics is that use temperature and humidity sensors data fusion for detecting fire [2, 9], and transportation surveillance by employing visual sensor network (VSN) [3]. Considered the complicacy as the municipal security is off, maximize relevancy multimedia sensors data should be involved in our case as well as exclude redundancy. That means strong correlation multimedia sensor data is necessary to effectively redeploy a multi-view and multi-sense monitor system, which consists of serial vision sensors out the field also the sonar, temperature and onto sensors etc. in the field.

The second, the study of complex networks plays an increasingly important role in information fusion. Examples of such networks are geographic grids and traffic networks, social relations, collaboration and citation networks of scientists, etc. The structure of such networks affects their performance. One main feature of complex networks is the fact that they are large. As a result, their complete description is utterly impossible, and researchers, both in the applications and in mathematics, have turned to their local description: How many vertices do they have, and by which local rules are vertices connected to one another? These local rules are probabilistic, which brings us to consider random graphs. These local rules are probabilistic, which brings us to consider random graphs.

In addition, the vision sensor network that affords successive video frames, conceal some correlational semantic knowledge transferred by series interactional target motion elements [4]. On the other side, by using the series multimedia sensors could detect lots of associated information, such as the noise caused by the crowd, the smell which may the clue of some toxic gas, even some detected suspicious words captured by the sound sensor. To overcome these problems, we will begin by looking at some valid models, which are suitable for video understanding and event detection. In this paper, we propose a new event recognition framework for consumer videos by leveraging a large amount of videos. As we know, graph structure provides a complex, dynamic, and robust framework for assembling complex relationships involved in the objects, which is suitable for our goal.

2. Related Works

Recent research on information fusion shows the expressive activity. Most of them concentrate on making the utmost use of the comprehensive sensor resources, which base on the inexpensive disposal of multiple sensors information. According to the diversity and complexity of the manifold sensors system, the method is required to robust and parallel processing ability. Furthermore, complementary and redundant information from different sensors which vary in the modalities and spatiality, asks for an optimizing description that a consistency of aimed objects.

The fusion of various modalities is generally performed at the feature level (premature fusion) and decision level (late fusion) [5]. The feature level fusion, extract the features from the original information referred by sensors, then analysis and process these extracted modalities information which may include visual feature(color, shape or texture involved), audio feature(frequencies or texture involved), motion features(motion trace or motion pattern involved) and so on. This approach achieves impressive compression also useful for real-time processing, which provides considerable feature information for the decision analysis, but it still disappoints in the time synchronization between the multimodal features [6]. The decision level fusion, on the other hand employs numerous sensors to observe one object, each sensor finishes own local processing (pretreatment, feature extraction, recognition and so on) then the insular decisions combine to a fused decision based on the correlation. This approach is advantageous in it utilizes unique representations for the diverse modalities features. Besides achieves great scalability in fusion process rather than the feature level, however it cost extra time caused by obtaining the limited decisions. Inherit the thought of the two levels; the author [7] introduces a recorder agent and a context agent: which respectively are that a specific camera with only recording features and indicate the semantic distance (monitor the same scene or target) between various agents, the former isolates the features process and the latter excludes the disparity in the modalities.

Base on diverse specific application background, it has plenty widely used fusion approaches involved in or concerned with relevant studies. In the application of current a part of them employ temperature, infrared and smoke sensors for environmental data

detection[2], and some utilize video sensor network(VSN) for intelligent transportation systems or surveillance systems [7], also information fusion is widely used in the field as robotics, image processing or some cross realms [8]. In paper [9], the author adopts the environmental data (temperature and humidity) as in-filed lower level data which a chosen thresholds can observe the change at each sensor node, then by using Dempster-Shafer Evidential reasoning(D-S) to fuse VSN data confirm the probability of fire. However, the supplying of the vision sensors is restricted to corroborate the assumption proposed as the first step; lots of associated video semantic knowledge were barely effective in timely warning then wasted. In additional by combining the Kalman filtering (KF) method with D-S [3], this thesis conquers the insufficient of both fusion algorithms and result in matching model, besides propose a strategy of camera topology estimation to prevent the probability of the overlap that occur in the reality, but also, the fusion of dynamic characteristics of the targets between successive video frames was still excluded in the theory. Moreover, devote to track and locate materials, information fusion model was adapted to the idea of automated identification and location estimation[10], the author provided a multi-level data fusion model based on the modified JDL model which joint D-S theory and weighted mean. The hybrid fusion method succeeded in fusing data from disparate sensor nodes that within noise and dynamic environment, besides advanced in the uncertainty and imprecision problem.

The way we propose basically inherits the thought of [7] in pretreating the modalities distance, whereas advanced dig the underlying connection between disparity sensor nodes that we attempt to adopt the strong correlation entre multi-sensor data. Accordingly redeploy the sensor data to adjust the environment, and then finish the fusion process at the decision level.

The field of random graphs was initiated in 1959-1960 by Erdos and Renyi. At first, the study of random graphs was utilized to prove deterministic properties of graphs. For example, if we can show that a random graph has a positive probability a certain property, then a graph must exist with this property. The method of proving deterministic statements using probabilistic arguments is called the probabilistic method, and goes back a long way. See among others the preface of a standard work in random graphs by Bollob as [12], or the work devoted to it on the Probabilistic Method [13]. All derive an analytical expression for the cluster coefficient, which shows that the graphs are distinctly different from standard random graphs, even for infinite dimensionality [14].

3. Undirected Graph and Multimedia Sensor Networks

As far as concerned, we will begin by looking at some minimal models of public security in crowds. Regardless of how different sensor data are organized, the underlying fusion algorithms must ultimately fuse the input data. Public security oriented fusion applications have to deal with several data related challenges. As a result, we chose to explore data fusion algorithms based on data-driven aspects of fusion. We aim to find the optimal sensing ranges of sensor nodes in a network with some targets. It is presumed that a sensor is able to dynamically adjust its sensing data by the arbitrary amount.

3.1. Definition of Events in Multimedia Sensor Networks

The automatic event detection was performed on the multimedia sensor networks to analyzing the spatial-temporal properties of individual calcium release events. It was composed by atomic events and composed events.

1.1.1. Atomic events: Firstly, we define *EV* as characteristic of *obj* in time *t*

$$EV(obj, t, feature) \tag{1}$$

On basis of this, we can describe the event as a set of scene from time k, length of the event is a series of scene, by use of set $I[k, k+\Delta k]$. If there exists relationships R1 and R2 between two scene $i \in I[k, k+\Delta k]$, $j \in I[k, k+\Delta k]$:

$$\begin{cases} R(T_i, T_j) \\ R2(EV(obj, T_i, feature), EV(obj, T_j, feature)) \end{cases} \tag{2}$$

R1 and R2 are used to describe the relationships among features in number. In order to understand atomic events more precisely, we employ definitions as below:

Definition 1: *State. The state* is an abstract of attributes owned by motion element, and is a static description of the condition and activity of a motion element at a certain time. *State* = {*Enter, Leave, Resort, ForwardAndBack* }, indicates the basic state of any motion element within the scope of spatial constraint in a georeferenced video stream, including the description information of *Appear*, *Disappear*, *Move* and *Stop*.

- *Enter*. The emerging motion element is newly enter and distinguished from the existing ones in the specific area of geographical boundary, and the state of which is called *Enter*. Then the motion element starts to be detected and tracked. *Enter* instance is regarded as the first instance of motion element.

- *Leave*. In contrast with the *Enter* state definition, *Leave* means the state of disappearance in the geographical boundary specific area or the untraceable state within a specific time, which is viewed as the last instance for the state description. *Leave* state is the signal of canceling motion element detection and tracking.

- *Resort*. *Resort R* is defined on triple $R = (Area(S), \zeta_{min}(S), \zeta_{max}(S))$. In detail, $Area(S)$ means the spatial plane area; $\zeta_{min}(S)$ and $\zeta_{max}(S)$ represent the maximum and minimum time threshold of *Resort* respectively. The *Resort* definition views the movement or stays that without markedly changed of space coordinate information within a certain region as motionless.

- *Move*. Within the scope of spatial constraint, *Move M* is a general designation of connecting the other three basic states in a continuous motion process of motion element. An instance of *Move* can be represented as $M = (Enter \mid Resort_k, Resort_{k+1} \mid Leave)$. By connecting the other three basic state instances, *Move* can form a linear sequence formed through the combination of *Enter*, *Leave* and *Resort*.

- *ForwardAndBack*. Within the scope of spatial constraint, *ForwardAndBack F* is a general designation of connecting the other three basic states in a continuous motion process of motion element. An instance of *Move* can be represented as $M = (Appear \mid Stop_k, Stop_{k+1} \mid Disappear)$. By connecting the other three basic state instances, *Move* can form a linear sequence formed through the combination of *Appear*, *Stop* and *Disappear*.

Definition 2: *Behavior Attribute*. Behavior description of a single typical motion element mainly includes spatial location and speed. Spatial location can be defined as $Location(Object) = (X_i, Y_i, T_i)$, which means that the spatial location of the motion element *Object* at time point T_i is (X_i, Y_i), X_i and Y_i represent the horizontal and vertical ordinate value in the two-dimensional plane respectively. $Speed(Object) = \{ S_{Value}, S_{Vector}, T_i \}$ indicates the motion element *Object* with velocity magnitude S_{Value} and velocity direction S_{Vector} at the time point T_i, among which S_{Vector} is the unit vector in a general planar domain.

Definition 3: *Relation*. *Relation* is an incidence relation of mutual influence between two motion element in the same time subspace T. *Relation* = (Object$_i$, Object$_j$, T) shows the relationship between motion element *Object$_i$* and *Object$_j$* in time subspace T which means one-dimensional time coordinates. The measurement of interaction established

between the two elements use probability *P*, which is dynamic adjustment with the influence of temporal-spatial factor, and $P \in [0, 1]$.

Definition 4: *Spatial Relation. Spatial Relation* includes measuring relation, direction relation and topological relation. *Spatial Relation SR = (Measure, Direction, Topology). Measure* indicates the measuring relation among motion element using some measure in measuring space, such as distance. In the same planar reference domain, *Direction* is the equity mutual relationship between source target and reference target.

1.1.2. Compound Event: Compound events is defined by the set of atomic events $E=\{e_1, e_2, \ldots, e_n\}$ obtained by event logic operation combination.

3.2 Construction of Undirected Graph

3.2.1. Undirected Graphs of Cluster Center: The random graph has the advantage that we can think of the graph as evolving as a process, by adding the edges one at a time, which also allows us to investigate dynamical properties. According to that, we connected each cluster center as a vertex, and take the Euclidean distance as the edge weight to form a weighted undirected graph *G=(V, E, W)* . In particular, V, E, W is the vertex set, edge set, and the edge weights real symmetric matrix, respectively. When the crowd motion in the scene is in a steady state for each clustering center position fluctuated in a small range. In addition, major changes of population movement have produced major change in the clustering center. Therefore, graph vertex space distribution reflects the distribution of velocity in dynamic scenes, and graph vertices represent clustering also implicitly contains location information to the class of people moving with speed.

The edge weight matrix *W* is the 2- norm distance between the vertexes. Therefore, the positive real matrix of *W* is $\{W_{i,j}\}^+$. $W_{i,j} = W_{i,j} \geq 0$, $W_{iij}=0$, $i, j, \in \{1,\ldots,N\}$, *N=length(V)* $\in Z^+$ is the number of vertices in G. Each vertex represents an average speed of all kinds of movement, and the distance between vertexes is the rate difference between the various types of movement. Therefore, the edge weight matrix describes the dynamic scene discrete clusters with varying degrees of speed. In order to fully examine the abnormal state of the crowd, speed and location features are not sufficient; we should also consider the density of all types of people, traffic and the size of other groups with similar speed information. β_i is the number of sample observations from the *i*th cluster, crowd density is the number of people and the area occupied by the provider, under the premise of a fixed scene, to simplify the calculations, the area occupied by all kinds of people unified area of constant usage scenarios square constant S, $\rho_i=\beta_i/S =\beta_i S'$. Therefore, we use the density information on i, j between two vertices of edge weights extended as:

$$\dot{W}_{ij} = (\rho_i + \rho_j)\dot{W}_{ij} = S'(\beta_i + \beta_j)\dot{W}_{ij} \qquad (4)$$

Therefore, the extended edge weight matrix *W* is $\{W'_{i,j}\}^+$, $W_{i,j}$ is the difference of two types of velocity, and the crowd flow is the product of crowd density and flow rate. Thus, formula 4 has the form of differential equations with flow $\rho(v^1 - v^2)$. Therefore, the dynamic changes of sequence W demonstrate the density and flow of the crowd.

3.3 Event Detection

As mentioned in the previous, edge weights contains a wealth of information about the structure and the state diagram, the sample size of the class where the vertex weighting coefficient is introduced as the edge weight, which may be more suitable to describe the change of scene, the crowd speed, density and flow. We define *n* as the rank of matrix *W*,

the Graph sequence $\{W\}^{\tau-1}$ Changes in the available dynamic systems in the time domain output, said:

$$\begin{cases} y(t+1) = Ay(t) \\ W(t) = Cy(t), t = 1,\ldots,\tau-1 \end{cases} \tag{5}$$

Y(t) is the hidden state variables, W is observations, $A \in R^{N \times n}$ is the implicit transition matrix, $C \in R^{N \times n}$ is the state transition matrix, and $N \geq n$。 We can judge the *state* by comparing the predicted and observed properties of the scene.

$$P^{-1}WP = \Lambda P^{-1} \Rightarrow W = P\Lambda P^{-1} \tag{6}$$

The outputs also satisfy a Markov property with respect to the dynamic model:

$$\begin{cases} \hat{C} = P \\ y(t) = \Lambda P^{-1}P \\ \hat{A} = Py(t+1)\,y(t)^{-1} \end{cases} \tag{7}$$

Take the formula (7) into equation (5), we can obtain the predicted value of the edge weight matrix W:

$$\hat{W}(t+1) = \hat{C}y(t+1) \tag{8}$$

Using Martin distance, we can get the deviation between W and $\hat{W}$:

$$D(\hat{W}, W) = \sqrt{\sum_{n=0}^{\infty} n \left| c_{\hat{W}}(n) + c_W(n) \right|^2} \tag{9}$$

If the crowd in a steady state, $D(\hat{W}, W) \propto 0$. On the contrary, if the crowd is in a unstable state, $D(\hat{W}, W)$ will increase quickly and suddenly.

Therefore, the state transition function of sparse random graph can be defined as a mapping relation.

$$\delta = \Theta \rightarrow \Theta' \tag{10}$$

Among them, Θ is the state space of sparse random graph, $\Theta=(D_1, D_2, \ldots, D_n)^{\mathrm{T}}$, and D is $D(\hat{W}, W)$.

The dynamic evolution process of sparse random graph includes its character update of motion element vertex V_N, emerging vertex with the *Enter* and *Leave* behavior state and the dynamic adjustment of edge set E and interaction relation P of random graphs. For the accurate description of event development process in georeferenced video stream, evolution rule algorithm of sparse random graph is shown as follows.

Algorithm 1: Evolution rule algorithm of sparse random graph.
Input: sparse random graph G_{Active}, motion element detection and recognition information;
Output: return G_{Active};
1. IF t = 0 Then
2.   Create first node S & Add S to V_N;
3. End IF
4. While t ≥ 1 do
5.   IF $V_{tmp} \rightarrow$ State Is Equal *Appear* Then
6.   Find nearest node V_{near};
7.   Create new edge $E(V_{tmp}, V_{near})$;
8.   Add Vt_{mp} to V_N;
9.   End IF
10.   For $V_i \in V_N$ do   //Update all Nodes in G_{Active}
11.   IF $V_i \rightarrow$ State IsEqual *Disappear* Then
12.   Remove V_i from V_N;
13.   Delete edge of V_i in G_{Active};
14.   End IF
15.   Update V_i;

16.   End For
17.   For $V_j \in V_N$ do   //Update all Edges in G_{Active}
18.   IF *Flag* ← getRestriction(V_j) Then
19.   Delete edge of V_j in G_{Active};
20.   Else IF *Flag* ← getAttract(V_j) Then
21.   Add new edge of V_j to G_{Active};
22.   End IF
23.   End For
24.   For $V_k \in V_N$ do   //Update P of Graph in G_{Active}
25.   IF $V_k \rightarrow$State IsEqual *Appear* Then
26. $P \leftarrow P(0)$;
27.   Else Update other P of G_{Active};
28.   End IF
29.   End For
30.   Return G_{Active};

4. Simulation and Analysis

While before the data fusion process, all the sensors information require for a unified expression, which means between the homogeneous or heterogeneous sensors can obtain decision from each other. In this case, we first quantize all the sensor data and transfer this quantized data to the information for decisions that acquire the semantics from various media information source. By employing the public monitor system, the quantized analysis of the correlation cohesion between motion elements can express the semantic knowledge underlying the video stream coming from each video sensor node. In our case, we fuse with the acoustic sensor, odor sensor data to detect the same scene from different respects. The Figure 1 depicts the fusion process and the multi-sensors data flow as following, and each modality node own exclusive judgments standard.
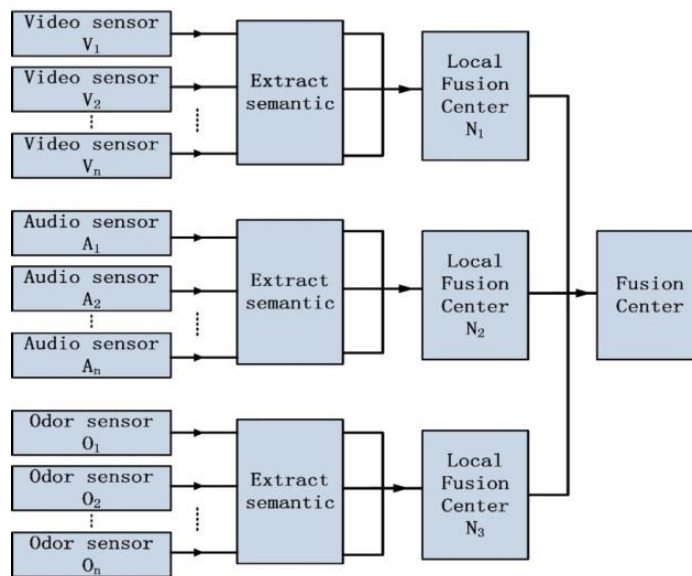

Figure 3. Fusion Process

Using the three scenes which involve some events that contain a group of people interact with each other, we carry out the comparison study. And all the chosen samples are considered as the labeled training data within the target domain.

Table 1. Abnormal Events in the Three Scenes

Abnormal Types	S_1	S_2	S_3
Cross-border	15	13	16
Flocking	17	17	17
Conflict	21	16	19

GBR accomplishes a concise numerical calculation and avoids the problems of computing complexity in the traditional CBR method. In Table 2, Table 3 and Table 4, we compare the performance of GBR with MKL [5] and Graph-Based Multi-Sensor Fusion method (GMF) in this paper using three different scenes.

Table 2. Comparisons of Different Methods in Scene 2

Method	Out-Dection	Correct-Dection	Omit-Dection
GBR	67	45	8
MKL	60	51	2
GMF	57	52	1

Table 3. Comparisons of Different Methods in Scene 2

Method	Out-Dection	Correct-Dection	Omit-Dection
GBR	58	38	7
MKL	51	43	2
GMF	52	45	0

Table 4. Comparisons of Different Methods in Scene 3

Method	Out-Dection	Correct-Dection	Omit-Dection
GBR	35	42	4
MKL	36	24	3
GMF	53	24	3

From Table 2, Table 3 and Table 4, we observe that GBR extends the processing time in a common detection of video event, but the forecasting accuracy of video abnormal behavior and emergency increased significantly with lower computation and complexity. Therefore, the energy consumption of sensors will be reduced which is consistency with the transmission costs, especially in the nonrecurring flocking emergency with complex video event modeling.

5. Conclusion

In this paper, we introduced a new sensor data centric method of heterogeneous multimedia sensors network, and explored associated frameworks and algorithms for incorporates useful contextual and temporal cues for public security alarm, leading to correlation analysis of the target motion elements in the multimedia sensor stream.

As a future work, we aim to advance implemented fusion algorithms integrated with alternative combination rules. In addition, attempt to deal with public events of alarm that are conducted to be optimized to multimedia sensed data and set appropriate fusion parameters as alarm thresholds, fusion weights and so on.

Acknowledgements

References

[1] G. B. Melton and G. Hripcsak, "Automated Detection of Adverse Events Using Natural Language Processing of Discharge Summaries", J Am Med Inform Assoc, 2005;12:448-457 doi:10.1197/jamia.M1794 .

[2] T. F. Castanedo, "A Review of Data Fusion Techniques," The Scientific World Journal, vol. 2013, Article ID 704504, 19 pages, (**2013**), doi:10.1155/2013/704504.

[3] A. Díaz-Ramírez and L. A. Tafoya, "Wireless Sensor Networks and Fusion Information Methods for Forest Fire Detection", Procedia Technology, (**2012**), pp. 69-79.

[4] Information Fusion and Its Application in Transportation Surveillance, A Dissertation for the Degree of Shanghai Jiao Tong University Philosophy Doctor

[5] J. Feng and H. Song, "Intelligent Analysis for Georeferenced Video Using Context-Based Random Graphs", International Journal of Distributed Sensor Networks, (**2013**).

[6] P. K. Atrey and M. A. Hossain, "Multimodal fusion for multimedia analysis: a survey", Multimedia Systems, (**2010**), pp. 345-379.

[7] Z. Wu, L. Cai and H. Meng, "Multi-level fusion of audio and visual features for speaker identification", International Conference on Advances in Biometrics, (**2006**), pp. 493-499.

[8] F. Castanedo and J. Garcia, "Data fusion to improve trajectory tracking in a Cooperative Surveillance MultiAgent Architecture", Information Fusion, (**2010**), pp. 243-255.

[9] X. Zhao, Q. Luo and B. Han, "Survey on Robot Multi-sensor Information fusion", Proceedings of the 7th World Congress on Intelligent Control and Automation.

[10] E. Zervas and A. Mpimpoudis, "Multisensor data fusion for fire detection", Information Fusion, (**2011**), pp. 150-159.

[11] S. N. Razavi and C. T. Haas, "Multisensor data fusion for on-site materials tracking in construction", Automation in Construction, (**2010**), pp. 1037–1046.

[12] B. Bollobas, "Random graphs, volume 73 of Cambridge Studies in Advanced Mathematics", Cambridge University Press, Cambridge, second edition, (**2001**).

[13] N. Alon and J. Spencer, "The probabilistic method. Wiley-Interscience Series in Discrete Mathematics and Optimization", John Wiley & Sons, New York, second edition, (**2000**).

[14] J. Dall Jesper and M. Christensen Michael, "Physical revie E, Statistical, nonlinear, and soft matter physics, (**2002**), 66: 016121.

Authors

JiangFan Feng, He received his B.S. degree from Southwest Agricultural University, and his Ph.D. degree from Nanjing Normal University, in 2002 and 2007. His main research area includes spatial information integration and multimedia geographical information system.

Mingbo Gou, She received the B.S. degree from QingHi Normal University in 2011. She is currently working towards her M.S. in the Chongqing University of Posts and Telecommunications. Her current research interests include Image and Video Analysis.

HaiBin Hu, he was born in 1980. He works as lecturer of China West Normal University. His main research area includes database and multimedia applications.