Research paper

A Variational Level Set Approach to Multiphase Multi-Object Tracking in Camera Network Base on Deep Features

E. Pazouki1*, M. Rahmati2
1Artificial Intelligence Department, Faculty of Computer Engineering, Shahid Rajaee Teacher Training University, Tehran, Iran.
2Artificial Intelligent and Robotics Department, Faculty of Computer and Information Technology Engineering, Amirkabir University of Technology, Tehran, Iran.

Article Info

Article History:
Received 02 September 2020
Reviewed 15 December 2020
Revised 25 January 2021
Accepted 01 March 2021

Keywords:
Multi-object tracking
Camera network tracking
MultiPhase level set representation
Variational tracking
Deep features

*Corresponding Author’s Email Address:
ehsan.pazouki@sru.ac.ir

Abstract

Background and Objectives: Object tracking in video streams is one of the issues in machine vision that has many applications. Depending on the type of the object, the number of objects and other inputs used in tracking, object tracking is divided into several different categories. Multi-object tracking in a camera network is one of the most complex types of object tracking. In this type of tracking, the goal of the algorithm is to extract the persistent trace of several objects moving simultaneously in a wide area that is monitored by a network of cameras. This type of tracking is often done in two steps. In the first step, the traces of each object in each camera is called tracklets are extracted. Then, the persistent trace of the objects are obtained by associating the extracted tracklets of all cameras in the monitored wide area. Here, we introduce a novel variational approach based on the deep features to associate the tracklets.

Methods: For this purpose a variational model with multiphase level set representation is introduced. The persistent trace of all objects are obtained by optimizing the proposed variational model. The proposed variational model is optimized by employing the Euler-Lagrange equation. CNN and deep learning are used to extract the deep features of appearance and motion of objects. Here, a ResNet50 network that is pre-trained on ImageNet and a transformer neural network which is trained with motion parameters of tracklets such as acceleration and orientation change rate are used for extracting deep features.

Results: The multiphase model using deep features presented in this paper provide 9% better results than the multiphase model without deep features based on TCF and FS metrics and 8% better results based on MT metric.

Conclusion: The results on the three well-known datasets which are real and a synthesized dataset show that the proposed model takes competitive performance, while using less extra context information of the camera network and objects, compared to the other proposed methods. The evaluations show the quality of the proposed model in solving complex problems using the minimum required initial knowledge.

©2021 JECEI. All rights reserved.
trajectories of objects presented and which are viewed by a camera network covering the area. The high level analysis is performed on the extracted trajectories and appropriate alarms are set. Multi-object tracking in a camera network is performed in two steps [1]. In the first step, moving objects are captured by all of the cameras within the view of camera network and are tracked by a tracking algorithms. Extensive research has been reported to deal with single camera tracking [2], [3]. There are various challenges exist for extracting accurate trace of objects in each camera, among them pose and illumination variations, occlusion, clutter and sensor noise are more common. Advancement of algorithms to resolve object tracking in a single camera have reached to a satisfactory extent [4], [5]. In this paper, the admissible trace of the objects in a single camera using the existing algorithms are extracted and are called tracklets. In the second step of multi-object tracking of a camera network, the extracted tracklets are associated to the corresponding objects and the persistent trace of objects are extracted. Usually, the association is performed based on the appearance of objects and motion models of the objects which are extracted from the tracklets. The solution space of the association task grows exponentially with the number of the tracklets [6]. Our proposed method provides a remedy for this problem. This problem is an ill-posed inverse problem [7]. The tracklets are the observation and are assumed to be known and the persistent trace is the ideal output and unknown. The variational model is an effective solution for solving the ill-posed inverse problem [8]. This model solves the ill-posed inverse problem in the image processing and computer vision [9], [10], [11] and [12]. In this paper, our main objective is to propose a variational multiphase model for solving the association problem.

Previous attempts on associating tracklets with corresponding objects place different restricted assumptions on the problem, thus deviating their application from real world problem [1]. In [13], [14] and [15], first order Markovian model is used for association. Using Markovian model for tracking objects in a single camera is a reasonable assumption but in a non-overlapped camera network is not a good choice where this assumption increases the probability of a wrong association. In some previous researches [14] and [16], the topology of cameras and the moving model of the objects are assumed to be known, while it is difficult to extract. Also, modeling the moving pattern and updating the changes are challenging.

Multi-object tracking in a camera networks is important issue. Associating of corresponding objects between different cameras distinguishes single camera and multi-camera tracking. Thus, the main challenge of the proposed model for tracking multi-object in camera networks is the association [1]. In some studies [17], [13], [16], [14], [18], and [6], the main goal is to associate the traces of objects and extract their persistent trace. In these researches, various tools and algorithms are used, including a planar tracking correspondence model (TCM), a Bayesian modeling, Parzen windows, the path coverage of a directed graph, a multi-objective optimization framework, a path smoothness, a statistical model function, a graphical representation and a network flow algorithm for modeling and association. In [19], an approach to improve the detection and tracking performance in multi-camera scenarios with overlapping field-of-views is provided, which allows for better handling of occlusion problem. It mainly includes monocular people detection, projection, fusion, probabilistic occupancy map generation and multi-object tracking steps. The objective of the study is to detect and track individuals within a designated open area where multiple cameras are set up, implement a robust multi-camera people detection and tracking method and improve the experimental performance.

With the introduction of deep learning [20] methods and the development of its applications in various issues, the use of models based on deep learning in objects tracking was also introduced in different methods. In some methods, the features of deep neural networks are used as visual or temporal features of objects [21] and [22]. For example in [21] the 2048-dim fully connected layer of ResNet50 [23] before the classification layer is used to represent the appearance of the objects. In some methods, models based on convolutional neural network (CNN) and deep learning are used to extract tracklets or persistent trace of the objects [22], [24], [25] and [26]. In [22], a CNN-based model called TrackletNet is introduced that uses the a graph to extract the trace of objects. In [24] a Long Short-Term Memory (LSTM) is used for tracking object. In [25] is provided a model based on Recurrent Neural Network(RNN) that models the changes made in the object by updating the features of the tracking object and performs better tracking of the object. In [26] a CNN-based model called Siam R-CNN is proposed which combine a Siamese re-detection architecture with a tracklet-based dynamic programming algorithm. In [27] a tracklet processing algorithm cleave and re-connect tracklets on crowd or long term occlusion which uses Siamese Bi-Gated Recurrent Unit (GRU). In this study, the tracklet is generated using deep features which are extracted by CNN and RNN to create the high-confidence tracklet candidates in sparse scenario. This proposed neural network model is trained using a dataset which contains more than 95160 pedestrian images. In [28] a multi-object tracking framework called DROP (Deep Re-identification Occlusion Processing) is
A lightweight convolutional neural network that can solve the re-tracking problem is constructed by increasing and learning the affinity of appearance features of the same object in different frames. In this study, the occlusion of the object is judged using the data association result of the appearance features of the object, and to reduce the matching error by improving the data association formula.

In some studies, the main objective is to introduce compositional features that are more effective in recognizing objects. In [29], a procedure is introduced which, using the classic features PHOG and CS-LBP and combining them with deep features and also using a new feature selection method such as JékNN, proposes an efficient combined feature for classifying objects. This combined feature is able to significantly improve results relative to similar models. In [30], deep features are combined with multiview features and a set of features is obtained for recognizing human behavior, which has good results compared to similar models. In order to model the objects with the aim of properly classifying them in [31], the combination of deep features, Very Deep Convolutional Networks for Large-Scale Image Recognition and Inception V3, are used. The obtained combined features are able to provide good sustainable recognition rate in object classification. Also in [32], by combining the deep features obtained from the two deep networks, VGG and AlexNet, with SIFT which is the classical feature for object detection, a combined feature has been created with the help of Réyni entropy-controlled method, which provides good results in classifying objects.

The combined features, some of which were reviewed in the above studies, are not practical in object tracking applications due to the considerable computational complexity involved in extracting them, as in object tracking the features of all objects in the image must be calculated for each frame. If they are used, the response time of the tracking algorithm will be significantly increased, which is not tolerable and acceptable in many real problems.

Another method of tracking objects is to use variational model, examples of which are presented in [33] and [34].

In [33], a variational model called single phase variational is introduced which is tracked in multi-objects in multi-camera network of wide area surveillance system. In [34], a novel variational model called multi phase variational model is proposed which is used the RGB color histogram as appearance model of the objects and the acceleration and orientation change rate as the motion model of the objects.

In this paper, we propose a novel deep variational method for associating the tracklets. In this method, the obtained method for associating the tracklets. In this method, the acceleration and orientation change rate of each object is represented as a multiphase level set function. By solving this association problem with less restrictive assumption, the optimum solution is reachable and the method is more general and usable in realistic scenarios. CNN and deep learning are used to extract the features of appearance and motion of objects.

In this paper, we propose a multiphase variational model for associating tracklets based on deep features. The proposed model is a variational optimization model that is converted to an Ordinary Differential Equation (ODE) which is solved numerically for extracting the persistent trace of objects.

The structure of the paper is as follows. In Section “Variational Model,” the proposed variational model and solving method are presented. The experimental results of the proposed model on the real and synthesized datasets are given in Section “Experimental Results” and finally, the conclusions are presented in Section “Conclusion.”
model as follow,
\[
J[R_p] = \int_{r_{p_t} \in R_p} \left( \lambda_1 \times \int_{T_t \in T_C} \text{CLS}_{r_{p_t}}(T_t) dl + \lambda_2 \times \int_{T_t \in T_C} \text{SM}_{r_{p_t}}(T_t) dl \right) d\tau
\]
where \( \text{CLS}_{r_{p_t}}(.) \) is closeness part of the variational model, \( \text{SM}_{r_{p_t}}(.) \) is smoothness part of the model and \( \lambda_1 > 0 \) and \( \lambda_2 > 0 \) are positive constants. The closeness is defined as, 
\[
\{ \begin{array}{ll}
T_t \in r_{p_t} & \text{if } \varphi_x(l) \geq 0 \\
T_t \notin r_{p_t} & \text{if } \varphi_x(l) < 0 
\end{array}
\]
where persistent trace of each object \( r_{p_t} \) is presented as a level set function \( \varphi_x \) and persistent trace of all tracked objects are presented as \( \phi = \{ \varphi_1, ..., \varphi_b \} \) which is a multiphase level set function. As stated in previous section the number of the objects which move in the wide area \( n \) is unknown and the number of tracked objects \( b \) is not necessarily equal to \( n \). Now, the (2) can be rewritten as,
\[
J[\phi] = \int_{r_{t=1}}^{\{|\phi|\}} \left( \lambda_1 \times \int_{l=1}^{\{|C|\}} \text{CLS}_{\varphi_x}(l) dl + \lambda_2 \times \int_{l=1}^{\{|C|\}} \text{SM}_{\varphi_x}(l) dl \right) H(\varphi_x(l)) d\tau
\]
where \( H(\varphi_x(l)) \) is the Heaviside function and is defined as,
\[
H(Z) = \begin{cases} 
1 & Z \geq 0 \\
0 & Z < 0 
\end{cases}
\]
The closeness and smoothness equations are redefined based on the level set representation, and are provided in [34].

C. Optimizing the Energy Function
In order to solve the persistent trace problem, the presented level set representation model must be optimized.
For optimizing this model, the Euler-Lagrange equation is used [36]. It must be noted that the regularized versions of the function \( H(z) \), denoted here by \( H_c(z) \), is used. The regularized version of (4) is defined as,
\[
J_c[\phi] = \int_{r_{t=1}}^{\{|\phi|\}} \lambda_1 \times \int_{l=1}^{\{|C|\}} \text{CLS}_{\varphi_x}(l) dl + \lambda_2 \times \int_{l=1}^{\{|C|\}} \text{SM}_{\varphi_x}(l) dl \right) d\tau
\]
where regularized version of the closeness and smoothness part of the model are redefined in [34].
In this paper, the regularized version of \( H(z) \) which is proposed in [37] is used. this function is defined as,
\[
H_c(z) = \frac{1}{2} \left( 1 + \frac{2}{\pi} \arctan \left( \frac{z}{2} \right) \right)
\]
The persistent trace of the objects is obtained by solving following optimization problem,
\[
\phi^* = \arg\min_{\phi} J_c[\phi]
\]
This optimization problem is solved by fixing, \( F_{A_c}, F_{M_c}, \text{Mean}F_{A_c} \) and \( \text{Mean}F_{M_c} \) then \( J_c[\phi] \) is minimized with respect to \( \phi \) using Euler-Lagrange equation [7].

B. Multi Phase Level Set Representation
Proposing the level set representation of the model starts with representing persistent trace of the objects as,
\[
\phi^* = \arg\min_{\phi} J_c[\phi]
\]
Then, the equation \( \varphi_L(t, l) \) is optimized. The numerical procedure required to solve (8) is presented in [34].

### Experimental Results

The performance of our proposed model is evaluated by performing several experiments using three challenging real datasets and one synthesized complex dataset.

The quality of the results of our experiments are determined based on well-known metrics used in camera network applications. The datasets and the metrics that are used are introduced in the following subsections.

#### A. Datasets

In this paper four challenging datasets are used. Three of them are real video sequences and another is synthesized data which is developed in our Lab.

First, the CAVIAR dataset [38] and [39] is collected in a shopping mall corridor with two cameras.

Second, the NGSIM dataset [40] is captured from Peachtree street located in Atlanta, Georgia by using eight synchronized cameras.

Third, the PETS2009 dataset [41] was collected through 8 cameras which are set up to monitor a road corner of a university campus. But, in this paper only four cameras are used.

In order to generate thorough annotated dataset according to every complicated surveillance scenario based on the given information a tool has been developed [42] for synthesing virtual data.

Here, a synthesized dataset with 6 cameras and 10 objects is used as forth experimental dataset.

The characteristics of these dataset are presented in Table 1.

The datasets’ details and the way they are used in this paper are consistent with the experiments performed in [34]. In order to show the status of the datasets, their figures are provided follow.

In Fig. 1 the sample image of one camera of the CAVIAR is presented and some samples of objects in different pose of this dataset are shown in Fig. 2.

In Fig. 3 five sample image of five different cameras of the NGSIM dataset are shown.

In Fig. 4 the sample images of two cars for 3 different pose are presented.

In Fig. 5 the sample images of the four different cameras of PETS2009 dataset are shown and the images of two different objects from different poses are shown in Fig. 6.

In Fig. 7 the world plane image of a wide area covered by 6 cameras from the simulator is presented.

In Fig. 8 some sample images of the objects from simulator are shown.
C. Results and Discussion

For evaluating the performance of the proposed model, the results are compared with four similar models in Table 3 and Table 4. As provided in Table 3 the proposed model tracks objects in wide area with the
average TCF metric of 78.8% and FS metric of 79.26% which means it extracts more than 79% of the objects’ persistent trace. These results show that the multiphase model using deep features presented in this paper can provide 9% better results than the multiphase model without deep features [34].

Also, the proposed model tracks objects with average MT metric of 85.60% which means it tracks more than 80% of the persistent trace of more than 85% of the objects. As a result, the deep features in terms of MT metric improve by an average of 8% compared to the other features [34]. In order to better compare the results of the proposed model with other models, two charts are presented in Fig. 9 and Fig. 10.

Figure 9 presents a chart for showing the effect of using the variational model and deep features simultaneously compared to the variational based models without deep features. The numbers in this chart are obtained by averaging the results obtained from the entire datasets. As can be seen in this figure, the proposed model in most metrics has provided better results than both variational models without deep features.

Also, in Fig. 10 the proposed model is compared with the results of the two other methods that are not variational based, which show the better performance of the proposed model. In other words, the proposed model presents competitive results compare to the similar models.

In Fig. 11, the computed persistent trace of one of the objects of NGSIM dataset and ground truth persistent trace of this object in the world plane are shown. As illustrated in this figure, the object are tracked in three cameras of the camera network. In Fig. 12, the persistent trace result of an object of PETS2009 dataset in two cameras is presented. Also, the extracted persistent trace of an object of synthesized dataset which has been tracked in four cameras of the camera network is given in Fig. 13.

<table>
<thead>
<tr>
<th>Name</th>
<th>Wide Area Width(meter)</th>
<th>Wide Area Height(meter)</th>
<th>#Cameras</th>
<th>#Tracklets</th>
<th>#objects</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAVIAR</td>
<td>30</td>
<td>65</td>
<td>1</td>
<td>413</td>
<td>88</td>
</tr>
<tr>
<td>NGSIM</td>
<td>150</td>
<td>650</td>
<td>5</td>
<td>691</td>
<td>195</td>
</tr>
<tr>
<td>PETS</td>
<td>50</td>
<td>55</td>
<td>4</td>
<td>53</td>
<td>10</td>
</tr>
<tr>
<td>Synthesis</td>
<td>50</td>
<td>50</td>
<td>6</td>
<td>102</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 2: The characteristics of the Metrics

<table>
<thead>
<tr>
<th>Name</th>
<th>Abbreviation</th>
<th>Unit</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Track Completion Factor [18]</td>
<td>TCF</td>
<td>Percent</td>
<td>0%</td>
<td>100%</td>
<td>Max</td>
</tr>
<tr>
<td>Track Fragmentation [18]</td>
<td>TF</td>
<td>Numerical#</td>
<td>1</td>
<td>-</td>
<td>Min</td>
</tr>
<tr>
<td>Physical Object ID Fragmentation [43]</td>
<td>POIF</td>
<td>Numerical#</td>
<td>0</td>
<td>1</td>
<td>Max</td>
</tr>
<tr>
<td>Precision [43]</td>
<td>PT</td>
<td>Percent</td>
<td>0%</td>
<td>100%</td>
<td>Max</td>
</tr>
<tr>
<td>Sensitivity [43]</td>
<td>ST</td>
<td>Percent</td>
<td>0%</td>
<td>100%</td>
<td>Max</td>
</tr>
<tr>
<td>F-Score [43]</td>
<td>FS</td>
<td>Percent</td>
<td>0%</td>
<td>100%</td>
<td>Max</td>
</tr>
<tr>
<td>ID Switching [44]</td>
<td>IDS</td>
<td>Numerical#</td>
<td>0</td>
<td>-</td>
<td>Min</td>
</tr>
<tr>
<td>Fragment [44]</td>
<td>FG</td>
<td>Numerical#</td>
<td>0</td>
<td>-</td>
<td>Min</td>
</tr>
<tr>
<td>Mostly Tracked [44]</td>
<td>MT</td>
<td>Percent</td>
<td>0%</td>
<td>100%</td>
<td>Max</td>
</tr>
<tr>
<td>Mostly Lost [44]</td>
<td>ML</td>
<td>Percent</td>
<td>0%</td>
<td>100%</td>
<td>Min</td>
</tr>
</tbody>
</table>
Table 3: The results of the proposed model

<table>
<thead>
<tr>
<th>Metrics</th>
<th>CAVIAR</th>
<th>NGSIM</th>
<th>PETS2009</th>
<th>Synthesized</th>
<th>CAVIAR</th>
<th>NGSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCF</td>
<td>73.14%</td>
<td>79.6%</td>
<td>77.26%</td>
<td>85.18%</td>
<td>-</td>
<td>67%</td>
</tr>
<tr>
<td>TF</td>
<td>1.52</td>
<td>1.18</td>
<td>1.16</td>
<td>1.28</td>
<td>-</td>
<td>1.39</td>
</tr>
<tr>
<td>POIF</td>
<td>0.38</td>
<td>0.34</td>
<td>0.48</td>
<td>0.35</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PT</td>
<td>84.21%</td>
<td>91.27%</td>
<td>76.12%</td>
<td>87.74%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ST</td>
<td>73.28%</td>
<td>70.11%</td>
<td>77.29%</td>
<td>78.15%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>FS</td>
<td>78.37%</td>
<td>79.3%</td>
<td>76.7%</td>
<td>82.67%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>IDS</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>-</td>
</tr>
<tr>
<td>FG</td>
<td>5</td>
<td>8</td>
<td>4</td>
<td>8</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td>MT</td>
<td>85.12%</td>
<td>89.39%</td>
<td>84.86%</td>
<td>83.02%</td>
<td>84.0%</td>
<td>-</td>
</tr>
<tr>
<td>ML</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>4.0%</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 4: The Results of Single Variational Model [33] and Multi Variational Model [34]

<table>
<thead>
<tr>
<th>Metrics</th>
<th>CAVIAR</th>
<th>NGSIM</th>
<th>PETS2009</th>
<th>Synthesized</th>
<th>CAVIAR</th>
<th>NGSIM</th>
<th>Single</th>
<th>Multi</th>
<th>Single</th>
<th>Multi</th>
<th>Single</th>
<th>Multi</th>
<th>Single</th>
<th>Multi</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCF</td>
<td>71%</td>
<td>71.97%</td>
<td>74%</td>
<td>76.3%</td>
<td>74%</td>
<td>75.52%</td>
<td>83%</td>
<td>64.13%</td>
<td>71%</td>
<td>72%</td>
<td>71%</td>
<td>72%</td>
<td>71%</td>
<td>72%</td>
</tr>
<tr>
<td>TF</td>
<td>1.61</td>
<td>2</td>
<td>1.21</td>
<td>1.27</td>
<td>2</td>
<td>1.22</td>
<td>2</td>
<td>2.5</td>
<td>1.61</td>
<td>2.28</td>
<td>1.61</td>
<td>2.35</td>
<td>1.61</td>
<td>2.40</td>
</tr>
<tr>
<td>POIF</td>
<td>0.37</td>
<td>0.31</td>
<td>0.31</td>
<td>0.32</td>
<td>0.46</td>
<td>0.41</td>
<td>0.33</td>
<td>0.28</td>
<td>0.37</td>
<td>0.28</td>
<td>0.37</td>
<td>0.28</td>
<td>0.37</td>
<td>0.28</td>
</tr>
<tr>
<td>PT</td>
<td>83.76%</td>
<td>79.68%</td>
<td>80.43%</td>
<td>89.39%</td>
<td>75.35%</td>
<td>70.04%</td>
<td>87.50%</td>
<td>76.52%</td>
<td>83.76%</td>
<td>76.85%</td>
<td>83.76%</td>
<td>77.62%</td>
<td>83.76%</td>
<td>78.85%</td>
</tr>
<tr>
<td>ST</td>
<td>72.58%</td>
<td>68.56%</td>
<td>69.83%</td>
<td>62.12%</td>
<td>76.14%</td>
<td>74.62%</td>
<td>77.62%</td>
<td>68.85%</td>
<td>72.58%</td>
<td>69.40%</td>
<td>72.58%</td>
<td>69.40%</td>
<td>72.58%</td>
<td>69.40%</td>
</tr>
<tr>
<td>FS</td>
<td>75.74%</td>
<td>71.33%</td>
<td>71.09%</td>
<td>69.69%</td>
<td>72.43%</td>
<td>68.79%</td>
<td>79.61%</td>
<td>69.40%</td>
<td>75.74%</td>
<td>70.32%</td>
<td>75.74%</td>
<td>71.09%</td>
<td>75.74%</td>
<td>71.09%</td>
</tr>
<tr>
<td>IDS</td>
<td>9</td>
<td>11</td>
<td>28</td>
<td>3</td>
<td>5</td>
<td>8</td>
<td>6</td>
<td>-</td>
<td>9</td>
<td>12</td>
<td>9</td>
<td>12</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>FG</td>
<td>6</td>
<td>7</td>
<td>35</td>
<td>9</td>
<td>8</td>
<td>10</td>
<td>12</td>
<td>-</td>
<td>6</td>
<td>12</td>
<td>6</td>
<td>12</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>MT</td>
<td>80.10%</td>
<td>78.32%</td>
<td>73.40%</td>
<td>87.09%</td>
<td>83.33%</td>
<td>70%</td>
<td>81.00%</td>
<td>75%</td>
<td>80.10%</td>
<td>78.32%</td>
<td>80.10%</td>
<td>78.32%</td>
<td>80.10%</td>
<td>78.32%</td>
</tr>
<tr>
<td>ML</td>
<td>2.00%</td>
<td>0%</td>
<td>12%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>2.00%</td>
<td>0%</td>
<td>2.00%</td>
<td>0%</td>
<td>2.00%</td>
<td>0%</td>
</tr>
</tbody>
</table>

Fig. 9: Graph comparing the results of the proposed model with two other variational models, including the single-phase model [33] and the multi-phase model [34] without using deep features.
A Variational Level Set Approach to Multiphase Multi-Object Tracking in Camera Network Based on Deep Features

Fig. 10: Graph comparing the results of the proposed model with the other two models.

Fig. 11: The persistent tracking results of the NGSIM for one object: a) Extracted result; and b) Ground truth result.

Fig. 12: The persistent tracking results of the PETS2009 for one object: a) Extracted result; and b) Ground truth result.
Conclusion

In this paper, we proposed a variational multiphase model for associating the tracklets of the objects in the camera network and determining their persistent trace. We proposed a new representation of the multi-object tracking problem in a camera network which can solve this problem with less restricted assumption. So, this model is a more general model for multi-object tracking in camera network which doesn't need the serious prerequisite information of the wide area, the camera topology and the objects' models. We use the deep feature for appearance and motion representation of objects.

We have evaluated our proposed model by four complicated datasets using 10 well known and common metrics. These evaluations show the quality of the proposed model in solving complex problems using the minimum required initial knowledge.

Author Contributions

Ehsan Pazouki designed and simulated, carried out the data analysis and collected the data and Mohammad Rahmati interpreted the results and wrote the manuscript.

Acknowledgment

The authors gratefully thank the anonymous reviewers and the editor of JECEI for their useful comments and suggestions.

Conflict of Interest

The author declares that there is no conflict of interests regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancy have been completely observed by the authors.

Abbreviations

- $C_i$: The $i$th Camera
- $C_{Mi}$: Calibration Parameters of the $i$th Camera
- $P_r$: The $r$th Object
- $t_s$: Start tracking time windows
- $t_e$: End tracking time windows
- $T_C$: Tracklets of all Cameras
- $T_{0i}$: Tracklets of the $i$th camera
- $R_P$: Persistent trace of all objects
- $r_{Px}$: Persistent trace of the $x$th object
- $n$: Count of moving objects
- $b$: Count of persistent tracked objects
- $\phi$: Multiphase Level Set representation of $R_P$
- $\varphi_r$: Level Set representation of $r_{P_r}$
- $\mid \cdot \mid$: The cardinal of the set
A Variational Level Set Approach to Multiphase Multi-Object Tracking in Camera Network Base on Deep Features

References


Biographies

Ehsan Pazouki is a Professor in the school of Computer Engineering at Shahid Rajaee Teacher Training University where he has been a faculty member since 2016. Ehsan completed his Ph.D. and M.S. at Amirkabir University. His research interests lie in the area of wide area surveillance, ranging from theory to design to implementation. He has collaborated actively with researchers in several other disciplines of computer science, particularly Cognitive Science. Ehsan has experience in various industries related to his specialization for more than 10 years. For additional information see https://www.sru.ac.ir/en/school-of-computer/ehsan-pazouki/

Mohammad Rahmati received the M.Sc. degree in electrical engineering from the University of New Orleans, in 1987 and the Ph.D. degree in electrical and computer engineering from the University of Kentucky, Lexington, Kentucky, in 1994. He is currently an associate professor in the Computer Engineering Department, Amirkabir University of Technology (Tehran Polytechnic). His research interests include the fields of pattern recognition, image processing, bioinformatics, video processing, and data mining. He is the chair of the department and he is also a member of IEEE Signal Processing Society.

Copyrights
©2021 The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, as long as the original authors and source are cited. No permission is required from the authors or the publishers.

How to cite this paper:
DOI: 10.22061/JECEI.2021.7649.417
URL: https://jecei.sru.ac.ir/article_1542.html