A STRUCTURED LEARNING-BASED GRAPH MATCHING FOR DYNAMIC MULTIPLE OBJECT TRACKING

*Dayu Zheng, Hongkai Xiong**

Department of Electronic Engineering
Shanghai Jiao Tong University
Shanghai 200240, P.R. China

Yuan F. Zheng

Department of Electrical and Computer Engineering
The Ohio State University
Columbus, OH 43210 USA

ABSTRACT

To correctly detect dynamic targets and obtain a record of the trajectories of identical targets in appearance over time, has become significantly more challenging and infers countless applications in biomedicine. In this paper, we propose a novel structured learning-based graph matching algorithm to track a variable number of interacting objects in dynamic environments. Different from previous approaches, the proposed method takes full advantage of neighboring relationships as edge feature in the structured graph. The target problem is regarded as structured node and edge matching between graphs generated from successive frames. In essence, it is formulated as the maximum weighted bipartite matching problem which is solved by dynamic Hungarian algorithm. The parameters of the structured graph matching model can be acquired in a stochastic graduated learning step in different dynamic environments. The extensive experiments on dynamic cell and football sequences demonstrate that the resulting approach deals effectively with complicated target interactions.

Index Terms— Multiple object tracking, structure feature, learning-based graph matching, dynamic environments.

1. INTRODUCTION

This work is concerned with the problem of tracking interacting objects in complicated dynamic environments. The difficulty of the dynamic tracking problem grows considerably with the increasing density of image objects, and the problem becomes a challenging issue due to frequent dynamic interactions between the objects. Extensive research on generic video object tracking can be categorized into two major classes: object representation and localization, e.g. mean-shift tracking [1], and filtering and data association, e.g. particle filtering [2]. The former is considered as local search with low computation and little information of motion and structure factors. The latter solves the target problem by sequentially estimating the state of object using a sequence of noisy measurements about the object states, and its existing variants fail to deal with the interaction of the object motions in structured environments. Therefore, there has been active research on learning based methods for analyzing and understanding behavior prediction in videos [3]. Along the depth of observations, this paper is motivated to handle complicated interactions of the targets in dynamic environments, which could involve the entry, exit, splitting, and touching occurrences.

*The work has been partially supported by the NSFC grants No. 60632040, No. 60772099, No. 60928003.

To detect dynamic multiple targets correctly and obtain a record of the trajectories of identical targets in appearance over time, sequential inference has recently been developed [4]. Instead of assuming the one-to-one mapping between observations and targets in traditional multiple hypothesis trackers, MCMC-based sequential tracking methods allow multiple temporal associations between observations and targets, and simulate the distribution of the association probability with a number of targets [5]. However, an excessively large number of samples would be required to approximate the underlying density functions with desired accuracy. As abstract representations for complex scenes, attributed graph matching problems could be formulated to find the close-to-optimum solution [6] where two graphs are considered isomorphic only if the correspondence between their vertices pairs up vertices with equal labels. To avoid the time-consuming manual labeling of correspondences, learning the parameters for graph matching has recently been concerned in practice [7]. With the insight of learning-based graph matching [8], this paper integrates the structural quadratic compatibilities on mutual association and local compatibilities on point pattern matching into the objective function to find the optimal assignment in a dynamic behavioral model.

In a first contribution of this paper, structure feature is presented to track a variable number of interacting objects in complicated dynamic environments. The proposed structure feature involves neighboring relationships including the lengths and angles of the edges in the structured graph, and represents nonlocal structure information of the whole graph. As a second contribution, the structured learning-based graph matching model is established, replacing the generic graph matching cost with a novel structured graph matching cost by incorporating the structured factor. The proposed structured graph matching problem is regarded as maximizing the matching cost of subgraphs that consist of structured nodes and edges. The parameters of the model can be acquired in a stochastic graduated learning step. Therefore, the multiple object tracking problem is actually considered as the maximum weighted bipartite matching problem, which can be solved by dynamic Hungarian algorithm. The tracking results of the scenes with complicated target interactions are demonstrated by extensive experiments on dynamic cell and football sequences, and the proposed approach achieves a good per-

978-1-4577-1303-3/11/$26.00 ©2011 IEEE

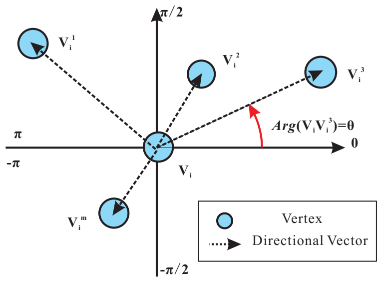


Fig. 1. The subgraph of the node and its neighborhood.

formance on dynamic object interactions.

The rest of the paper is organized as follows. The proposed structure feature is described in section 2. Section 3 demonstrates the structured learning-based graph matching method. The experimental results in Section 4 is validated to reflect the effectiveness of the proposed method for tracking a variable number of interacting objects.

2. STRUCTURE FEATURE

In generic graph matching methods, basic features for each separate target are recorded by a node, which would be utilized in node matching and parameter acquisition. However, it can not obtain a good performance by using the node features only, due to making little use of the whole structure in the graph and ignoring the environment related factors, especially in complicated dynamic environments. In this paper, the structure of the subgraph is the neighboring relationship between the objects, which is illustrated in Fig. 1. In essence, the probability of correct matching would be much larger when all the object and its neighborhood can match the co-located subgraphs in the next graph. Hence, we make inference with the structure feature in the graphical model.

In the sequence of images, graphs generated from two consecutive images are denoted by G and G', respectively. Each graph is a complete set of all nodes and edges (i.e., $G = (V, E)$) derived from the target image. Each detected object is defined as a node in the graph. Node set generated from image G with elements k is expressed by:

$$V = \{v_1, v_2, v_3, \cdots, v_k\}$$

To define structure feature, we need the concept of neighborhood. The neighborhood set of a target node consists of all neighbors adjacent to it, denoted by:

$$N(v_i) = \{v_i^1, v_i^2, v_i^3, \cdots, v_i^m\}$$

where m is size of the neighborhood set.

The size of the neighborhood set, or number of its neighbors, is also called as degree of a node, which can be decided in several different ways. In this paper, we choose to construct regulated graphs in which all nodes are endowed with the same degree value.

The structure feature is derived from edge feature which is the relationship between the nodes, including the lengths and angles of the edges between the object and its neighborhood. The edge feature is not only related with the node feature of the object's neighborhood, but the separation angle between two matching edges also makes a great effect. The angular factor considers degree values between two edges $e(v_i^{\alpha} v_i)$ and $e(v_{i'}^{\beta} v_{i'})$. In this way, the edge feature is calculated by:

$$\begin{aligned} & f_e(v_i^{\alpha}, v_i, v_{i'}^{\beta}, v_{i'}) \\ = \; & F_c(v_i^{\alpha}, v_{i'}^{\beta}) * exp\Big(- \frac{|Dist(v_i^{\alpha} v_i) - Dist(v_{i'}^{\beta} v_{i'})|}{dist} \Big) \\ & * exp\Big(- \frac{|Arg(v_i^{\alpha} v_i) - Arg(v_{i'}^{\beta} v_{i'})|}{\theta} \Big) \end{aligned}$$

where $v_i^{\alpha} \in N(v_i)$, $v_{i'}^{\beta} \in N(v_{i'})$, the function $f_e(\bullet)$ is the similarity of edge pair, $F_c(\bullet)$ is the similarity of node pair, and $Dist(\bullet)$ and $Arg(\bullet)$ are the length and angle of the edge respectively.

The subgraph consists of the object and its neighborhood as a part of the whole graph. The proposed structure feature involves neighboring relationships in dynamic environments, and represents nonlocal structure information of subgraphs. Therefore, the dynamic targets can be tracked easily by using the structure feature. In the proposed method, we propose a algorithm not simply in term of a node matching, but a subgraph matching method instead. The structure feature is also fully utilized in computing the matching cost to improve the tracking performance. Finally, the total structure feature $F_e(\bullet)$ of the subgraph can be defined:

$$F_e(v_i, v_{i'}) = \sum_{v_i^{\alpha_k} \in N(v_i), v_{i'}^{\beta_k} \in N(v_{i'})} f_e(v_i^{\alpha_k}, v_i, v_{i'}^{\beta_k}, v_{i'})$$

where $v_i^{\alpha_k} \in N(v_i)$ and $v_{i'}^{\beta_k} \in N(v_{i'})$, $k = 1, 2, \cdots, m$.

3. STRUCTURED LEARNING-BASED GRAPH MATCHING

3.1. Structured Graph Matching Problem

In the structured graph matching problem, we denote the notations used in the model, and define the structured graph matching problem. We denote a pair of graphs in a typical way, the first one G and the second G', and G_i as the i^{th} attribute of the node and G_{ij} as the edge ij in graph G. In the standard graphs, the edge attributes $G_{ij} \in \{0, 1\}$ are binary.

For the matching matrix y, $y_{ii'} \in \{0, 1\}$, $y_{ii'} = 1$ if node i in the first graph matches node i' in the second one, and $y_{ii'} = 0$ otherwise. $c_{ii'}$ is defined as the coefficient of the compatibility function for linear assignment ($i \to i'$), and $d_{ii'jj'}$ is defined as the coefficient of the compatibility function for quadratic assignment ($ij \to i'j'$). The graph matching problem is formulated in a generic way. Therefore, the estimated matching matrix $\hat{y}$ is acquired by the solution of the following optimization problem.

$$\begin{aligned} \hat{y} = \arg\max_y & \left[\sum_{ii'} c_{ii'} y_{ii'} + \sum_{ii'jj'} d_{ii'jj'} y_{ii'} y_{jj'} \right] \\ \text{s.t.} \quad & \sum_i y_{ii'} \leq 1, \text{for all } i' \\ & \sum_{i'} y_{ii'} \leq 1, \text{for all } i \end{aligned} \tag{1}$$

In this problem, there are some special constraints of the edges. The edges only exist between the objects and their neighborhood, which can be formulated as follows.

$$\begin{cases} d_{ii'jj'} \neq 0, \; v_j \in N(v_i) \; and \; v_{j'} \in N(v_{i'}) \\ d_{ii'jj'} = 0, \; otherwise \end{cases}$$

In this case, the matching cost in Eq. (1) is formulated as

$$\begin{aligned} & \textstyle\sum_{ii'} c_{ii'} y_{ii'} + \sum_{ii'jj'} d_{ii'jj'} y_{ii'} y_{jj'} \\ = & \textstyle\sum_{ii'} y_{ii'} (c_{ii'} + \sum_{v_j \in N(v_i), v_{j'} \in N(v_{i'})} d_{ii'jj'} y_{jj'}) \\ = & \textstyle\sum_{ii'} y_{ii'} (c_{ii'} + d'_{ii'}) \end{aligned} \tag{2}$$

3.2. Structured Learning-based Graph Matching

In the proposed approach, the parameters of the structured graph matching model are acquired in a structured learning phase. The training dataset is N observations x from an input set $\mathcal{X}$ and N corresponding labels from an output set $\mathcal{Y}$, which make up the structured training pairs of dataset $\{(x^1, y^1), (x^2, y^2), \ldots, (x^N, y^N)\}$, and x^n is an observation of graphs G^n and G'^n, and y^n is the n-th iterative match matrix between G^n and G'^n. Therefore, the learning problem becomes finding a parameterized function of the graph matching model g_ω, which minimizes the matching cost on the test dataset.

$$\frac{1}{N} \sum_{n=1}^{N} \Delta(g_\omega(G^n, G'^n), y^n) + \lambda \Omega(\omega) \tag{3}$$

where $\Delta(g_\omega(G^n, G'^n), y^n)$ is the loss incurred by the predictor g when predicting. The output $g_\omega(G^n, G'^n)$ is the prediction of the matching matrix y^n, which is used instead of the training input y^n in the method. The term $\Omega(\omega)$ (i.e., a regularization function of ω) and λ is a parameter in the loss, which are used against overfitting in the training dataset.

Next, we have to define the function $g_w(G, G')$, whose parameter ω we optimize over the loss function Δ and the regularization term $\Omega(\omega)$. In order to specify the function $g_\omega(G, G')$, we use the standard approach of discriminant functions. The discriminant function $f(G, G', y; \omega)$ is maximal for the case of $g_w(G, G')$, which is the optimal estimate for y (i.e., $g_\omega(G, G') = \arg\max_f(G, G', y; \omega)$). In a generic way, we define $f(G, G', y; \omega)$ as a linear function, $f(G, G', y; \omega) = \langle \omega, \Phi(G, G', y) \rangle$. Correspondingly, the predictor $g_w(G, G')$ is formulated in Eq. (4).

$$g_\omega(G, G') = \arg\max_{y \in \mathcal{Y}} \langle \omega, \Phi(G, G', y) \rangle \tag{4}$$

The joint feature of graph pairs have to be defined in order to specify the function $g_\omega(G, G')$, and the feature should contain the properties of both graphs as well as the properties of a match matrix y between these graphs. For this purpose, we can find the relationships between the learning phase given by Eq. (4) and the graph matching model previously given by Eq. (2), and the solution of the optimization problem of graph matching is the estimate of function g, i.e., $y^\omega = g_\omega(G, G')$. The discriminant function in Eq. (4) is introduced into Eq. (2):

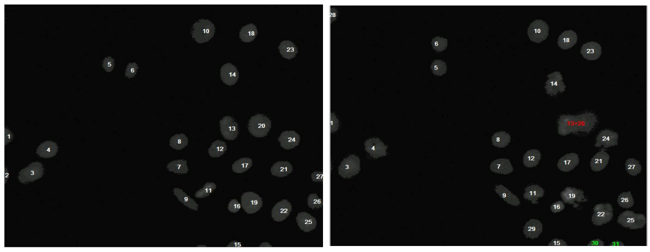


Fig. 2. Example tracking results obtained for cell sequence.

$$\langle \Phi(G, G', y), \omega \rangle = \sum_{ii'} y_{ii'} (c_{ii'} + d'_{ii'}) \tag{5}$$

The graphs and the parameters must be encoded in the compatibility functions. Like $f(G, G', y; \omega)$, we choose the coefficients of the compatibility functions also in linear method as:

$$\begin{aligned} c_{ii'} &= \langle F_c(v_i, v'_{i'}), \omega_1 \rangle \\ d'_{ii'} &= \langle F_e(v_i, v'_{i'}), \omega_2 \rangle \end{aligned} \tag{6}$$

where $F_c(v_i, v'_{i'})$ represents the node matching cost of node pairs $(v_i, v'_{i'})$, and $F_e(v_i, v'_{i'})$ is the edge matching cost in the subgraph pairs of the nodes v_i and $v'_{i'}$. In the learning phase, the appropriate feature can be selected arbitrarily. v_i and $v_{i'}$ are a potential candidate pair $\{v_i, v_{i'}\}$ that $v_i \in G, v_{i'} \in G'$.

In the particular case of Eq. (5), by defining $\omega := [\omega_1 \; \omega_2]$, $c_{ii'}$ and $d_{ii'jj'}$ are only related with the features of node and edge of graphs. We obtain the final form of $\Phi(G, G', y)$ from Eq. (5) and Eq. (6):

$$\Phi(G, G', y) = [\sum_{ii'} y_{ii'} F_c(v_i, v'_{i'}), \sum_{ii'} y_{ii'} F_e(v_i, v'_{i'})] \tag{7}$$

At last, we define the loss function

$$\Delta(y, y^n) = 1 - \frac{1}{\|y^n\|^2} \sum_{ii'} y_{ii'} y^n_{ii'}$$

and $\Omega(\omega)$ is specified as $\frac{1}{2}\|\omega\|^2$.

In the proposed method, the convex function $\frac{1}{N}\sum_n \xi_n$ is an upper bound for $\frac{1}{N}\sum_n \Delta(g_\omega(G^n, G'^n), y^n)$ with appropriately chosen constraints. The optimization problem of structured learning-based graph matching becomes

$$\begin{aligned} & \min_{\omega, \xi} \tfrac{1}{N} \textstyle\sum_{n=1}^{N} \xi_n + \tfrac{\lambda}{2}\|\omega\|^2 \\ \text{s.t.} \quad & \langle \omega, \Psi^n(y) \rangle \geq \Delta(y, y^n) - \xi_n \\ & \text{for all } n \text{ and } y \in \mathcal{Y}. \end{aligned} \tag{8}$$

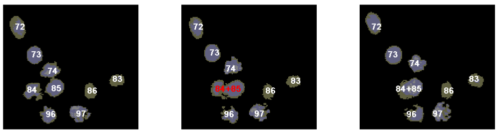
where $\Psi^n(y) = \Phi(G^n, G'^n, y^n) - \Phi(G^n, G'^n, y)$.

Referring to [8], we can acquire ω in a graduated learning step, and find the results of the structured graph matching problem in Eq. (1) using dynamic Hungarian algorithm.

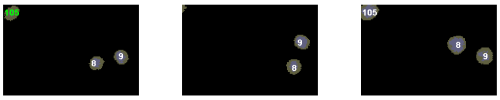
4. EXPERIMENTAL RESULTS

To validate the efficiency of the method, the proposed structured learning-based graph matching method has been applied to typical scenes.

(a) Touching Cells



(b) Merging Cells



(c) Boundary Cells

Fig. 3. Tracking results in dynamic cell environments.

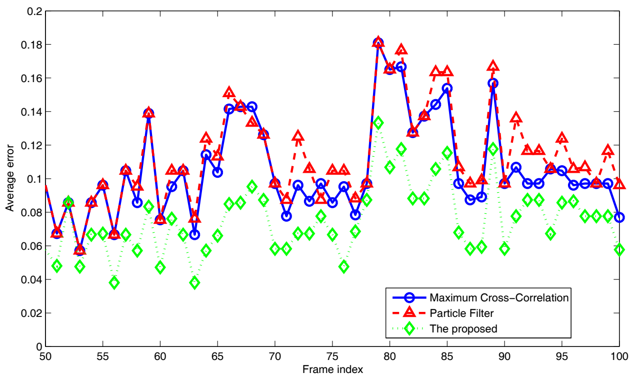


Fig. 4. Statistical performance in different methods

In the first experiment, we consider the difficult cell tracking problem in biomedicine. The cell sequence consists of 100 frames sampled from time-lapse fluorescence microscopy. These image sources are captured in a spatial resolution of 672×512, and a temporal resolution of 3 minutes between every two consecutive frames. Furthermore, the sources are full of different kinds of complicated cell interactions. According to the low intensity of the cell sequence, we utilize the LBF level set method for cell segmentation. The node features adopt the spatial, gray level, shape, and the proposed structure feature. The tracking results are shown in Fig. 2, and Fig. 3 illustrates the proposed method has a good performance for touching, merging and boundary scenes. Also, we use the Maximum Cross-Correlation and particle filter methods for comparison. Fig. 4 provides the average errors on each frame, which show the proposed structured learning-based graph matching method always gains a better performance on dealing with the complicated target interactions.

Another experiment on the football sequence is 2010 UEFA Champion League final, where we use SIFT and the proposed structure feature. The tracking results are shown in Fig. 5, where players are labeled in red rectangle, and the blue lines record the trajectories of players. It is illustrated that the proposed method performs well on the dynamic football sequence.

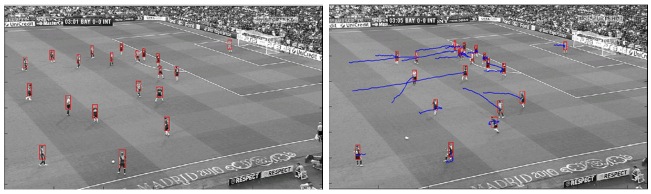


Fig. 5. Example tracking results obtained for football sequence.

5. CONCLUSIONS

To address the dynamic scenes in multiple object tracking problem, a novel structured learning-based graph matching method is proposed in this paper. In the proposed method, we utilize both the node and structure feature in the graphs, instead of considering the node feature only. The structured learning-based graph matching model is established using a novel structured matching cost by incorporating the proposed structure feature, and the parameters of the model are acquired in a stochastic graduated structured learning step. At last, the extensive experiments on cell and football sequences validate that the proposed method can gain a good performance in different scenes, especially in complicated dynamic environments.

6. REFERENCES

[1] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel based object tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 5, pp. 564-577, May 2003.

[2] C. Hue, J.-P. Le Cadre, and P. Perez, "Sequential monte carlo methods for multiple target tracking and data fusion," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 309-325, Feb. 2002.

[3] W. Hu, X. Xiao, Z. Fu, D. Xie, T. Tan, and S. Maybank, "A system for learning statistical motion patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1450-1464, Sept. 2006.

[4] Z. Khan, T. Balch, and F. Dellaert, "MCMC-based particle filtering for tracking a variable number of interacting targets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 11, pp. 1805-1819, Nov. 2005.

[5] K. Smith, D. Gatica-Perez, and J.-M. Odobez, "Using particles to track varying numbers of interacting people," In *CVPR*, pp. 962-969, 2005.

[6] T. S. Caetano, T. Caelli, D. Schuurmans, and D. A. C. Barone, "Graphical models and point pattern matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1646-1663, Oct. 2006.

[7] M. Leordeanu and M. Hebert, "Unsupervised learning for graph matching," In *CVPR*, pp. 864-871, 2009.

[8] T.S. Caetano, J.J. McAuley, Li Cheng, Q.V. Le, and A.J. Smola, "Learning graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 6, pp. 1048-1058, Jun. 2009.