# DETERMINING TOPOLOGY IN A DISTRIBUTED CAMERA NETWORK

*Xiaotao Zou, Bir Bhanu, Bi Song,* and *Amit K. Roy-Chowdhury*

Center for Research in Intelligent Systems, University of California, Riverside, CA 92521, USA

**ABSTRACT**

Recently, 'entry/exit' events of objects in the field-of-views of cameras were used to learn the topology of the camera network. The integration of object appearance was also proposed to employ the visual information provided by the imaging sensors. A problem with these methods is the lack of robustness to appearance changes. This paper integrates face recognition in the statistical model to better estimate the correspondence in the time-varying network. The statistical dependence between the entry and exit nodes indicates the connectivity and traffic patterns of the camera network, which are represented by a weighted directed graph and transition time distributions. A nine-camera network with 25 nodes is analyzed both in simulation and in real-life experiments, and compared with the previous approaches.

***Index Terms—*** camera network, topology, statistical model

## 1. INTRODUCTION

Networks of video cameras are being envisioned for a variety of applications and many such systems are being installed. However, most existing systems do little more than transmit the data to a central station where it is analyzed, usually with significant human intervention. As the number of cameras grows, it is becoming humanly impossible to analyze dozens of video feeds effectively. Therefore, we need methods that can automatically analyze the video sequences collected by a network of cameras.

Most work in image processing has concentrated on a single or a few cameras. While these techniques will be useful in a networked environment, more is needed to analyze the activity patterns that evolve over long periods of time and large swaths of space. Recently, there has been some work on understanding the topology of a network of non-overlapping cameras [1][2][3][5] and using this to infer about activities viewed by this network [6]. The authors in these papers proposed an interesting approach for modeling activities in a camera network. They defined either the entry/exit points in each camera or single cameras as nodes and learned the connectivity between these nodes. This provided an understanding of the paths that can be followed by objects within the field of view of the network of cameras. We build upon these ideas to develop a method for learning the network topology in an unsupervised manner by integrating identity and appearance information. The paper does not deal with how to optimally place these cameras; it focuses on how to infer the connectivity given fixed locations of the cameras. We now highlight the relation with the existing work and the main contributions of this paper along these lines.

## 2. RELATED WORK AND CONTRIBUTIONS

There is a lot of work on camera calibration and the topology inference of camera networks under the assumption of known data correspondence, which is not always guaranteed in the real-life environment. With respect to the increasing use of non-overlapping cameras in distributed camera networks, there is the need for new methods to relax the assumption. Makris *et al.* [1] proposed to use the temporal correlation of departures (*i.e.*, 'exit') and arrivals (*i.e.*, 'entry') to infer the network topology with unknown correspondence. Kieu *et al.* [3] used the information theoretic-based statistical dependence to infer the camera network topology. They proposed to integrate out the uncertain correspondence using Monte Carlo Markov Chain method. Marinakis *et al.* [5] used the Monte Carlo Expectation-Maximization (MC-EM) algorithm to simultaneously solve the data correspondence and network topology inference problems.

All these approaches take only the discrete departure and arrival events as input. To employ the abundant visual information provided by the imaging sensors, Niu and Grimson [2] proposed an appearance-integrated cross-correlation model for topology inference on the vehicle tracking data. However, appearances may be deceiving when the objects in the applications are humans. For example, clothing of different subjects is similar (*e.g.,* Fig. 1(a) and (c)), or appearance changes of the same subject under different illumination may be significant (*e.g.*, Fig. 1(a) and (b)).

In Fig. 1, the clothing of the subjects is similar, and the

(a) *A* in camera *1* (b) *A* in camera *2* (c) *B* in camera *2*

**Figure 1.** An example of false appearance similarity information. two subjects ('*A*' and '*B*') are monitored by two cameras ('*1*' and '*2*').

illumination of the two cameras is different. The Bhattacharyya distances between the RGB color histograms of the extracted subjects in the above three frames ('a,' 'b,' and 'c') are calculated to identify the subjects: $d$(a,b)=0.9097; $d$(a,c)=0.6828, which will establish a false correspondence between 'a' and 'c.'

Therefore, we propose a principled approach to integrate the appearance and identity (*e.g.*, face) to enhance the statistical dependence estimation for topology inference.

The main contribution of the paper is: *integrating appearance and identity for learning network topology*. The work in [2] uses the similarity in appearance to find correlations between the observed sequences at different nodes. However, appearances may be deceiving in many applications as in Fig. 1. For this purpose, we integrate human identity (*e.g.*, face recognition in our experiments) whenever possible in order to learn the connectivity between the nodes. We provide a principled approach for doing this by using the joint statistical model of appearance and identity to weight the cross-correlation. We show thorough simulation and real-life experiment results about how adding identity can improve the performance significantly over existing methods.

## 3. TECHNICAL APPROACH

In this section, we will show how to determine the camera network topology by measuring the statistical dependence of the nodes with the appearance and identity. The proposed approach to network topology inference is illustrated in the block diagram in Fig. 2.

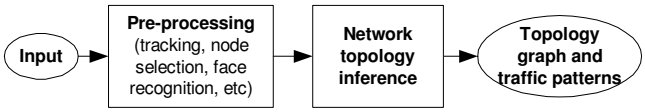


**Figure 2**. The block diagram of the proposed method.

### 3.1. Camera Network Topology

The nodes in our network architecture denote the entry/exit points in the field-of-views (FOVs) of all cameras in the network as in [1]. They can be manually chosen or automatically selected by clustering the ends of object trajectories [1]. If they are in the same FOV or in the overlapped FOVs, it is easy to infer the connectivity between them by checking object trajectories through the views. In this paper, we focus on the inference of connectivity between nodes in non-overlapping FOVs, which is blind to the cameras. The camera network topology is represented by a weighted, directed graph with nodes as entry/exit points and the links indicating the connectivity between nodes (as shown in Fig. 3).

Suppose we are checking the existence of the link from node $i$ to node $j$. We observe subjects departing at node $i$ and arriving at node $j$. The departure and arrival events are represented as temporal sequences $X_i(t)$ and $Y_j(t)$, respectively. We define $A_{X,i}(t)$ and $A_{Y,j}(t)$ as the observed

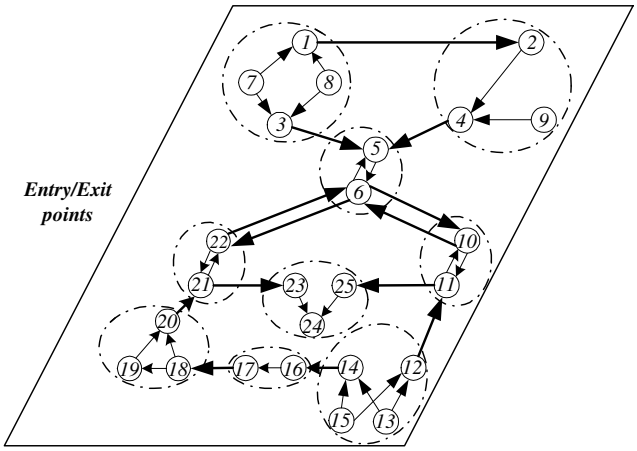


**Figure 3.** Topology graph of the deployed camera network.

appearances in the departure and arrival sequences, respectively. The identities of the subjects observed at the departure node $i$ and at the arrival node $j$ are $I_{X,i}(t)$ and $I_{Y,j}(t)$, respectively.

To alleviate the sole dependence on appearance, which may be deceiving as shown in Fig. 1, we propose to use the appearance and identity information together to weight the statistical dependence between different nodes, *i.e.*, the cross-correlation function of departure and arrival $X_i(t)$ and $Y_j(t)$:

$$R_{X,Y}(\tau) = E\left[X_i(t)\cdot Y_j(t+\tau)\right] = \sum_{t=-\infty}^{\infty} X_i(t)\cdot Y_j(t+\tau)$$

$$= E\left[f\left(A_{X,i}(t), A_{Y,j}(t+\tau), I_{X,i}(t), I_{Y,j}(t+\tau)\right)\right] \quad (1)$$

where $f$ is the statistical model of appearances and identity, which implicitly indicates the correspondence between subjects observed in different views. The components and joint models of $f$ are presented in the following sub-sections. From now on, we assume departure and arrival nodes are always $i$ and $j$, respectively, so that the subscripts $i$ and $j$ can be omitted.

### 3.2. Statistical Model of Identity

The working principle of the human identification is as follows: 1) detect the departure/arrival subjects and apply image enhancement techniques if needed (*e.g.*, the super-resolution method for face recognition); 2) the subjects departing from node $i$ are assigned unique identities $I_X(t)$ and used as the gallery; 3) the identities of the subjects arriving at the node $j$ (*i.e.*, $\tilde{I}_Y$) are verified by comparing it with all subjects in the gallery:

$$S_{ID}(\tilde{I}_Y) = \arg\max_{I_X}\left(sim(I_Y, I_X)\right) \quad (2)$$

where $sim(I_Y, I_X)$ is the similarity score between $I_Y$ and $I_X$, and $S_{ID}(.)$ is the highest similarity score associated with the identified identity.

We use a $k$-component mixture of Gaussian distribution (*e.g.*, as shown in Fig. 4) to model the similarity scores of identities ($S_{ID}$):

$$P_{ID} = P(S_{ID}(\tilde{I}_Y) \mid X = Y) = \sum_{m=1}^{k} a_m \cdot N(\mu_m, \sigma_m^2) \quad (3)$$

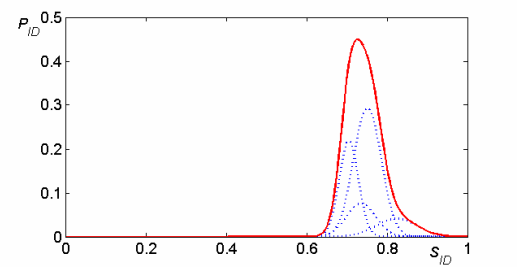


**Figure 4.** The Gaussian mixture model of the identity similarity.

where $k$ is the number of components, $\alpha_m$, $\mu_m$ and $\sigma_m^2$ are the weight, mean and variance of the $m^{th}$ Gaussian component, and '$X$=$Y$' means that they correspond to the same subject.

The unknown parameters $\{k, \alpha_m, \mu_m, \sigma_m^2\}$ can be estimated by using the EM algorithm [4] in face recognition experiments on a large data set. The mixture of Gaussians (as the solid line in Fig. 4), which has four components (as dotted lines in Fig. 4), is obtained by using the EM algorithm in the identification experiments [7].

### 3.3. Statistical Model of Appearance Similarity

We employ the comprehensive color normalization (as in [2]) to alleviate the dependence of appearances on the illumination condition. Then, the color histograms in the hue and saturation space, *i.e.*, $h$ and $s$, respectively, are calculated on the normalized appearance. Note that we do not incorporate the size information in the appearance metrics because the human subjects do not vary much in size. We normalize the sizes (*i.e.*, heights and widths) of the extracted subjects before calculating color histograms.

Next, a bi-variate Gaussian distribution is fitted to the color histogram similarity between the two appearances:

where $\mu_{h,s}$ and $\Sigma_{h,s}$ are the mean and covariance matrix of the color histogram similarity, which are learned by using the EM algorithm on the labeled training data.

$$P_{app} = P(h_X - h_Y, s_X - s_Y \mid X = Y) \sim N(\mu_{h,s}, \Sigma_{h,s}) \quad (4)$$

### 3.4. Joint Model of Identity and Appearance Similarity

By integrating the above statistical models of appearances and identity, the statistical model $f$ in Eq. 1 can be updated as the joint distribution of identity and appearance similarity, which are collectively denoted as $S=\{h_X - h_Y, s_X - s_Y, S_{ID}\}$:

$$\begin{aligned} & P_{similarity}\left(S \mid X(t), Y(t+\tau)\right) \\ & = P_{app}\left(X(t), Y(t+\tau)\right) \cdot P_{ID}\left(X(t), Y(t+\tau)\right) \\ & = P\left(h_X - h_Y, s_X - s_Y \mid X(t) = Y(t+\tau)\right) \cdot P\left(S_{ID}(\tilde{I}_Y) \mid X(t) = Y(t+\tau)\right) \end{aligned} \quad (5)$$

In Eq. 5, the joint distribution is the product of the marginal distributions of each under the assumption that the appearance and identity are statistically independent. For each possible node pair, there is an associated multi-variate mixture of Gaussians with unknown mean and variance, which are estimated by using the EM algorithm. We can even relax the independence assumption provided that we have enough training samples to learn the covariance matrix of the joint distribution.

Then, the cross-correlation function of departure and arrival sequences in Eq. 1 is updated as:

$$R_{X,Y}(\tau) = \sum_{t=0}^{\infty} P_{similarity}\left(S \mid X(t), Y(t+\tau)\right) \quad (6)$$

### 3.5. Network Topology Validation

The mutual information between two temporal sequences ([2]) reveals the dependence between them:

$$I(X,Y) = \int p(X,Y) \log \frac{p(X,Y)}{p(X) \cdot p(Y)} dXdY = -\frac{1}{2} \log_2 (1 - \rho_{X,Y}^2) \quad (7)$$

where $\rho_{X,Y}^2 \approx \dfrac{\max(R_{X,Y}) - \text{median}(R_{X,Y})}{\sigma_X \cdot \sigma_Y}$

Thus, we can use the mutual information as "threshold" to validate the existence of links identified in the previous cross-correlation model. The normalized mutual information is used as the weight of the links in the topology graph:

$$W_{i,j} = \frac{I_{i,j}(X,Y)}{M_I}, \quad \text{where } M_I = \arg\max_{(i,j)} \left(I_{i,j}(X,Y)\right) \quad (8)$$

## 4. EXPERIMENTAL RESULTS

We tested our proposed approach in simulation and in real-life experiments, and compared it with the appearance-integrated approach [2], when applicable.

### 4.1. Simulated Experiments

The simulation is based on the network architecture illustrated in Fig. 3. Since we focus on the connectivity inference in non-overlapping FOVs, the nodes with all connections within the same FOV are omitted. Thus, the simulated network has 18 departure/arrival nodes and 13 valid directed links. Some nodes, *e.g.*, node 11, work as both 'departure' and 'arrival.' Some node pairs, *e.g.*, 6 and 22, have two uni-directional links, which models the asymmetric traffic between the throughput nodes such as doors. The traffic data of 100 points is generated by a *Poisson*(0.1) departure process, and the transition time follows Gamma distributions, *e.g.*, *Gamma*(100, 5), *Gamma*(25, 2.5), etc. The probability of identity similarity $P_{ID}$ is generated by a mixture of Gaussians as shown in Fig. 4. For simplicity, the probability of appearance similarity $P_{app}$ is modeled by a uni-variate Gaussian $N(0, 1)$.

The proposed approach is tested on the simulated traffic data. We assume all the transition time distributions are single-mode. The cross-correlations with the appearance and identity (as in Eq. 6) for three valid and three invalid links are shown in Fig. 5(a) and (b), respectively. For comparison, Fig. 5(c) and (d) show the appearance-based cross-correlations [2] for the same valid and invalid links, respectively. It can be found that our approach can highlight the peaks for the valid links and repress fluctuations for the invalid links, which greatly improves the peak signal-to-noise ratios of the estimation.

As to the link validation, we calculate the mutual information of departure and arrival sequences at various nodes and show the adjacency matrices in Fig. 6(a, b). Based on the adjacency matrices, the topology graph is inferred as

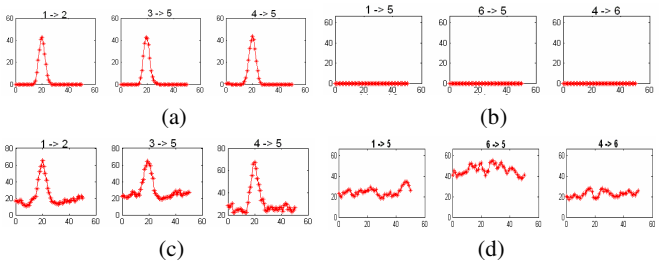


**Figure 5**. The estimated cross-correlations. (a, b) our proposed approach, (c, d) the previous approach in [2]. (a, c) are for valid links and (b, d) for invalid links.

shown in Fig. 6(c). In addition to the thirteen valid links (marked as solid arrows), the appearance-based approach [2] also generates nine invalid links (as the dashed arrows in Fig. 6(c)), which are mainly associated with the throughput nodes, *e.g.*, 11, 12 and 18.

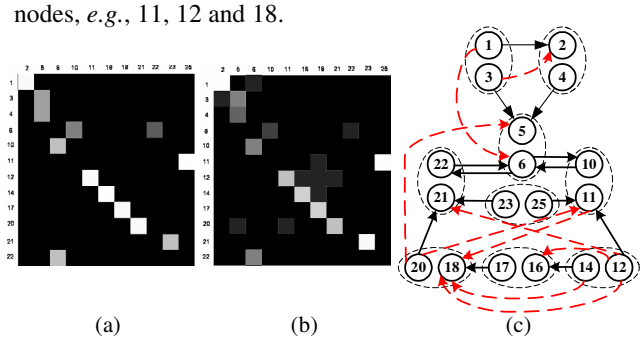


**Figure 6**. The adjacency matrices of mutual information: (a) by our approach; (b) by the previous approach in [2]; (c) the inferred topology graph.

### 4.2. Real-life Experimental Results

The experimental setup of the distributed camera network is illustrated in Fig. 7. As in the simulation, it has the same topology graph as shown in Fig. 3. Within it, there are nine cameras, in which six are on the tables (marked as circles) and three are on the ceiling (marked as triangles), distributed in two rooms on two floors. There are four doors monitored by four cameras, where the heavy traffic occurs. There are also some barriers in the rooms that constrain possible paths.

We collected data on a set of ten subjects: each person walked through the monitored environment ten times, totally 100 observations. The identification system is under construction so that we simulated the identity similarity distribution according to the mixture of Gaussians as in Fig. 4. After a manual selection of entry/exit points in each FOV (as ellipses in Fig. 7), the object detection and tracking were employed to detect the departure and arrival events. Subsequently, the probability of the appearance similarity was calculated as in sub-section 3.3, and the probability of the appearance similarity $P_{app}$ was calculated based on the estimated distribution from the labeled training data.

The proposed approach was tested on the real-life data to infer the network topology. It successfully recovered the topology of the camera network without any false link. However, the appearance-based approach [2] established several false links, to name a few, '2 to 6' and '4 to 16,' by accumulating false correspondences. For example, in Fig. 8,

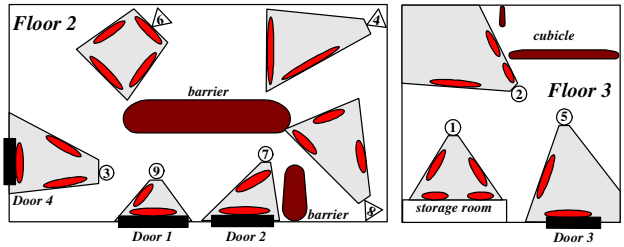


**Figure 7**. Experimental setup of the camera network showing the locations, FOVs, and entry/exit points of the cameras.

(a) and (b) are the same subject observed at nodes 16 and 6, respectively, and (c) is another subject at node 4. Their identities (*i.e.*, faces) are shown in the corner of each frame. Unfortunately, the false correspondences 'a=c' and 'b=c' are established by using the appearance similarity metrics. Therefore, the false links '4 to 6' and '4 to 16' are inferred by accumulating these false correspondences.


**Figure 8**. An example of false correspondence by appearance similarity metrics between different subjects. (a, b): one subject observed at nodes 16 and 6, respectively; (c): the other one at node 4.

## 5. CONCLUSIONS

Unlike the existing methods that used discrete events or appearance information to infer the network topology, this paper integrates identity with the appearance and provides statistical models to learn the dependences between nodes. Experiments in both simulation and real-life tests are performed that demonstrate the underlying proposed theory.

## ACKNOWLEDGEMENT

This work was partially supported by NSF awards CNS-0551741 and ECCS-0622176. The contents of the information do not necessarily reflect the position or policy of the U.S. government.

## 6. REFERENCES

[1] D. Makris, T. Ellis, and J. Black. "Bridging the gaps between cameras," In *CVPR,* Vol. 2: 205-210, 2004.

[2] C. Niu and E. Grimson. "Recovering non-overlapping network topology using far-field vehicle tracking data," In *ICPR*, Vol. 4: 944-949, 2006.

[3] K. Tieu, G. Dalley, and E. Grimson. "Inference of non-overlapping camera network topology by measuring statistical dependence," In *ICCV*, pp. 1842-1849, 2005.

[4] A. Dempster, N. Laird, and D. Rubin. "Maximum likelihood from incomplete data via the EM algorithm," In *Journal of the Royal Statistical Society*, Series B, 39(1): 1-38, 1977.

[5] D. Marinakis, G. Dudek, and D. Fleet. "Learning sensor network topology through Monte Carlo Expectation Maximization," In *ICRA*, pp. 4581-4587, 2005.

[6] D. Makris and T. Ellis. "Learning semantic scene models from observing activity in visual surveillance," In *IEEE Trans. on SMC*, *Part B*, 35(3): 397-408, June 2005.

[7] R. Wang, and B. Bhanu. "Learning models for predicting recognition performance," In *ICCV*, pp. 1613-1618, 2005.