CAMERA LINK MODEL ESTIMATION IN A DISTRIBUTED CAMERA NETWORK BASED ON THE DETERMINISTIC ANNEALING AND THE BARRIER METHOD

Chun-Te Chu, Jenq-Neng Hwang
Department of Electrical Engineering, Box 352500
University of Washington, Seattle, WA 98195, USA
{ctchu, hwang}@u.washington.edu

Yi-Yuan Chen, Shen-Zheng Wang
Service Systems Technology Center,
Industrial Technology Research Institute
Hsinchu, Taiwan 31040, R.O.C
{yiyuan, st}@itri.org.tw

ABSTRACT

Distributed camera networks have been deployed in the modern surveillance systems. The camera link model, including transition time distribution and brightness transfer function that represent the space-time relationship and color model between two cameras, is a critical element for tracking objects across the cameras. In this paper, we formulate the estimation of the camera link model as an optimization problem, where the deterministic annealing and the barrier method are applied to effectively extract the model parameters. Through the unsupervised scheme, our method utilizes the time stamps and color features together to establish the camera link model in presence of the outliers. Several simulations and comparative studies show the effectiveness of our approach.

***Index Terms*—** distributed camera networks, camera link model, deterministic annealing, barrier method, multiple camera tracking

1. INTRODUCTION

Due to the limited field of view (FOV) of a single camera, a surveillance system nowadays consists of several cameras covering a range of area. However, tracking objects across the cameras suffers from the non-overlapping FOVs between cameras. Before performing the multiple cameras tracking [1][2][12], how to obtain the reliable camera link model from the training videos becomes a critical issue. This model includes the transition time distribution which describes the traveling time between two cameras [4], and the brightness transfer function (BTF) that stands for the mapping between color models in different cameras [3]. If we know the correct correspondence between the objects in two cameras given the training videos, a set of traveling time between the match pairs can be used to estimate the transition time distribution by applying kernel density estimation [2]. Similarly, the BTF can be extracted, too. However, as the scale of the camera network is getting larger, unsupervised learning is more feasible than supervised ones. Moreover, in practice, since the connections between cameras may be arbitrary, the outliers exist; that is, one departing from a camera does not necessarily enter the other camera. Here we formulate the estimation as an optimization problem and use an unsupervised learning scheme by applying the deterministic annealing and the barrier method to build the model in presence of the outliers.

Makris et al. [4] proposed an estimation method to build the transition time distribution based on the cross-correlation between the exit and entry time stamps of the objects. Their assumption on the single mode distribution is inadequate since it cannot represent most cases in the real world. In [5], an entropy-based method was presented to find the distribution. Markov Chain Monte Carlo (MCMC) was utilized to find the optimal correspondence. Although they discussed the outliers and tried to model the color deviation between cameras, the performance reliability remained unclear. Moreover, it has been shown in [7] that their minimum-entropy assumption may not hold for practical applications. In Gilbert's work [6], they used color information in building the transition time distribution. However, similar to [4], they considered all the correspondence within a given time window which generated the mixture of true and false correspondences; hence, large amount of training data is required for building a reliable model. Also, they only dealt with the transition time distribution during the estimation process, and the BTF was handled separately. Huang et al. [7] introduced a method based on the Gibbs sampling, where they used the hard decision to determine the correspondence during the whole estimation process, which may be easily trapped in the local optimum. Furthermore, they did not take into account the BTF and the outlier issue.

Our aim in this paper is to present an unsupervised method that effectively estimates the camera link model given the training data containing the outliers. More specifically, (1) We formulate the camera link model estimation as an optimization problem and apply the deterministic annealing combined with the barrier method to find the optimal solution. (2) The time stamp and color information are both incorporated in the objective function that enables the simultaneous extraction of the transition time distribution and the BTF. (3) A reliable model is built based on this unsupervised learning scheme even in presence of the outliers.

This paper is organized as follows: Section 2 provides the optimization problem formulation of the camera link model. The estimation process is described in Section 3. Section 4 shows the experimental results, followed by the conclusion in Section 5.

2. PROBLEM FORMULATION

To construct the camera link model between the cameras, the correspondence between two observation sets from a pair of cameras need to be identified first given the training data. Inspired by the concept of feature points matching between two images [8], our modeling can be formulated as an optimization problem. Assume we have two sets of observations, $\mathbf{X}$ and $\mathbf{Y}$, which represent the exit and entry observations in two cameras.

$$\mathbf{X} = [\mathbf{x}_1 \quad \cdots \quad \mathbf{x}_{N_1}],\ \mathbf{Y} = [\mathbf{y}_1 \quad \cdots \quad \mathbf{y}_{N_2}]$$

where $\mathbf{x}_i$ and $\mathbf{y}_i$ are d-dimension feature vectors of an exit or entry observation, and N_1 and N_2 are the numbers of the observations in each set. Our goal is to find the $(N_1+1)\times(N_2+1)$ permutation matrix $\mathbf{P}$. Each entry P_{ij} in $\mathbf{P}$ will be set to 1 if $\mathbf{x}_i$ corresponds to $\mathbf{y}_j$; otherwise, it is set to 0. The $(N_1+1)th$ row and $(N_2+1)th$ column represent the outliers entries. Note P_{N_1+1,N_2+1} has no physical meaning, so all the following discussion will exclude it automatically. Hence, the problem can be written as a constrained minimization integer programming problem:

$$\mathbf{P}^* = \arg\min_{\mathbf{P}} J(\mathbf{X}, \mathbf{Y}, \mathbf{P}) \tag{1}$$

$$s.t. \quad P_{ij} \in \{0,1\} \quad \forall\, i \le N_1+1,\ j \le N_2+1 \tag{2}$$

$$\sum_{i=1}^{N_1+1} P_{ij} = 1 \quad \forall\, j \le N_2 \tag{3}$$

$$\sum_{j=1}^{N_2+1} P_{ij} = 1 \quad \forall\, i \le N_1 \tag{4}$$

where J is the objective function to be minimized. The constraint equations (2)~(4) enforce the one-to-one correspondence (except the outliers). The problem can be relaxed by substituting constraint (2) with

$$P_{ij} \ge 0 \quad \forall i \le N_1+1,\ j \le N_2+1 \tag{5}$$

In this way, the variable is continuous and can be easier to solve [8]. Moreover, the relaxation will reduce the chance of getting trapped in the local minimum. It can be proved that as the iteration proceeds, the solution will converge at the one to the original integer problem [9]. We then apply the deterministic annealing combined with the barrier method to iteratively solve the problem.

3. CAMERA LINK MODEL ESTIMATION

The objective function is divided into several parts as the following:

3.1. Transition Time Constraint

In order to build the transition time distribution f_T by using kernel density estimation, a set of time values $\mathbf{T}$ need to be collected first. Given the current estimation of $\mathbf{P}$, we can extract the transition time values $\mathbf{T} = [t_1 \quad \cdots \quad t_{N_1}]$ from $\mathbf{P}$, $\mathbf{X}$ and $\mathbf{Y}$:

$$t_i = \sum_{j=1}^{N_2} P_{ij} y_j^t + P_{i,N_2+1} x_i^t - x_i^t \qquad \forall i = 1 \sim N_1 \tag{6}$$

where P_{ij} can be used to indicate how likely the matching is between the *i-th* exit and the *j-th* entry observations; y_j^t and x_i^t are the entries representing the time stamps in the observation vectors $\mathbf{y}_\mathrm{j}$ and $\mathbf{x}_\mathrm{i}$. The transition time is always positive if two cameras have no overlapping area, i.e., entry time of a person is greater than the exit time of the correct correspondence. Also, if the *i-th* exit observation is an outlier, P_{i,N_2+1} should be one resulting in $t_i = 0$. Hence, we express this constraint ($t_i \ge 0$) in terms of a barrier function [11]:

$$-\frac{1}{\beta}\sum_{i=1}^{N_1} \log t_i \tag{7}$$

where β is a factor that effectively control the degree of the constraint satisfaction.

3.2. Transition Time Distribution

Given a set of time values $\mathbf{T} = [t_1 \quad \cdots \quad t_{N_1}]$, the estimation of the transition time distribution f_T is built based on the kernel density estimation:

$$f_T(t) = \frac{1}{N_1 \sigma_T \sqrt{2\pi}} \sum_{i=1}^{N_1} \exp\left(-\frac{(t-t_i)^2}{2\sigma_T^2}\right) \tag{8}$$

where σ_T^2 is the predefined variance of the Gaussian kernel. The confidence of the estimation is based on a maximum likelihood approach, i.e., for each possible correspondence, we compute the log likelihood value given the model. Thus, the total cost can be written as:

$$-\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} P_{ij} \log(f_T(y_j^t - x_i^t)) \tag{9}$$

3.3. Brightness Transfer Function (BTF)

Same object may appear differently in two disjoint cameras due to the illumination change and different camera responses, and the color deviation can be modeled as the brightness transfer function (BTF) [3]. The BTF is applied to compensate the color difference between two cameras before we calculate the distance between the histograms of two observations from two cameras. Thus, the total cost function for the BTF is:

$$\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} P_{ij} D(f_{BTF}(\mathbf{y}_j^h), \mathbf{x}_i^h) \tag{10}$$

where D is the distance function between two histograms; $\mathbf{y}_\mathrm{j}^\mathrm{h}$ and $\mathbf{x}_\mathrm{i}^\mathrm{h}$ are the color histograms of the observation vectors, $\mathbf{y}_\mathrm{j}$ and $\mathbf{x}_\mathrm{i}$; and f_{BTF} is the estimation of BTF. If two observations are actually a match, and the BTF is estimated correctly, the distance value will be small.

3.4. Maximum Entropy Principle

In deterministic annealing, a widely used scheme to solve the optimization problems, the procedure starts with emphasizing high "uncertainty", measured as the entropy, of the entries in $\mathbf{P}$; that is, to maximize the entropy. As the iteration goes, the importance of the maximum entropy principle is decreasing by controlling a certain parameter [10]. Thus, the cost function is written as the negative of the entropy:

$$\frac{1}{\gamma}\sum_{i=1}^{N_1+1}\sum_{j=1}^{N_2+1} P_{ij} \log P_{ij} \tag{11}$$

Note that P_{N_1+1,N_2+1} is not considered here. The factor γ starts with a low value that makes the importance higher in the beginning, and it then gradually increases to lower the emphasis. The function (11) can also be seen as the barrier function for the constraint (5); hence, we can set γ value the same as β in Section 3.1.

3.5. Objective Function

By combining the above cost functions, our final objective function to be minimized is shown in the following:

$$\begin{aligned} J(\mathbf{X},\mathbf{Y},\mathbf{P}) = &-\frac{1}{\beta}\sum_{i=1}^{N_1}\log t_i - \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} P_{ij}\log(f_T(y_j^t - x_i^t)) \\ &+ \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} P_{ij} D(f_{BTF}(\mathbf{y}_j^h), \mathbf{x}_i^h) + \frac{1}{\beta}\sum_{i=1}^{N_1+1}\sum_{j=1}^{N_2+1} P_{ij}\log P_{ij} \end{aligned} \tag{12}$$

Constraints (3) and (4) can be satisfied by performing the alternative row-column normalization based on the Sinkhorn's theorem [9]. By employing the deterministic annealing method and the barrier method [9][10][11], the $\mathbf{P}$ matrix is updated in each

```
Initialize P, f_T, f_BTF, β = β_0 .
While (β < β_f)
    P_ij = exp(−β ∂J/∂P_ij)
    Alternatively perform row-column normalization.
    Update f_T and f_BTF based on Section 3.6
    Update β = β × r
End
Update final f_T and f_BTF based on Section 3.6
```

Figure 1. Pseudo code of the optimization algorithm.

iteration with the decreasing of the objective function (12)(Fig. 1). In the early stage, due to the problem relaxation (see (5)), the P_{ij} value is continuous in the range between 0 and 1. Instead of making hard decision, this soft assignment prevents it from being trapped in the local minimum. The β value is also increased by a fix rate r until the final value β_f is reached. It can be proved that as the iteration increases, the $\mathbf{P}$ matrix eventually converges to a binary-valued matrix [9]. Since $\mathbf{P}$ determines the current estimation of the correspondence, we use it to update the camera link model.

3.6. Camera Link Model Update

For a current $\mathbf{P}$, the set of transition time $\mathbf{T} = [t_1 \quad \cdots \quad t_{N_1}]$ is extracted via (6). After that, the f_T is updated by applying kernel density estimation based on (8). To update the f_{BTF}, we should extract the corresponding histograms between two cameras. Given a current $\mathbf{P}$, each P_{ij} indicates how likely the matching is between the *i-th* exit and the *j-th* entry observations. Thus, we can calculate the weighted sum among the histograms in set $\mathbf{X}$ and $\mathbf{Y}$ separately, where the weights are set in proportional to P_{ij}.

$$\mathbf{h}_{\mathbf{X}} = \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} P_{ij}\mathbf{x}_i^h \qquad \mathbf{h}_{\mathbf{Y}} = \sum_{j=1}^{N_2}\sum_{i=1}^{N_1} P_{ij}\mathbf{y}_j^h \tag{13}$$

In this way, the histograms of the outliers will not be included in the cumulated histograms $\mathbf{h}_{\mathbf{X}}$ and $\mathbf{h}_{\mathbf{Y}}$. The f_{BTF} can thus be estimated and updated [3].

4. EXPERIMENTAL RESULTS

We have tested our proposed method in several scenarios. The observation sets are manually generated from the videos and the ground truth correspondence can also be identified. We follow the parameters setting in [9] by assigning $\beta_0 = 0.00091, r = 1.075$, and $\beta_f = 150$. σ_T in (8) is set as 0.5, and some other initial values are set as the following:

$$P_{ij} = \begin{cases} 0, & \mathbf{y}_j^t - \mathbf{x}_i^t < 0, \ \ \forall i = 1{\sim}N_1, j = 1{\sim}N_2 \\ e, & otherwise \text{ (include outlier row and column)} \end{cases}$$

where e is set as 0.01. The initial f_T is set as in Section 3.6 based on the initial $\mathbf{P}$, and the identity function is used as initial f_{BTF}. The Euclidean distance is employed for the distance function D in Section 3.3.

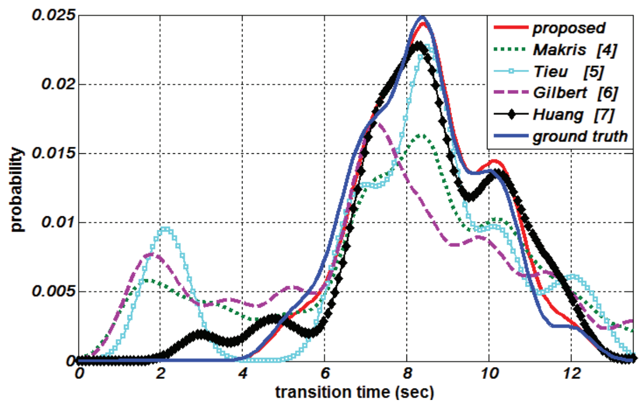

Figure 2. Transition time distribution for self-recorded video.

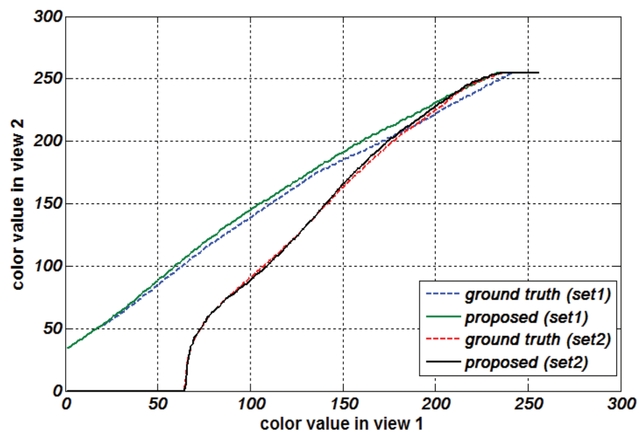

Figure 3. Brightness transfer function. The broken lines and the solid lines denote the ground truth and the results, respectively. Only one channel is shown here.

4.1. Set 1: Self-Recorded Data

Several cameras are mounted outside the department building with non-overlapping FOVs. Here we only show the simulations from one pair of cameras. There are total 80 people in the videos and contain 23% unmatched outliers. Fig. 2 shows the estimation results of the normalized transition time distribution. The results from other approaches [4,5,6,7] are also shown in different curves. Based on the results, it can be seen that our method produces the best match to the ground truth among all competing ones. Since other approaches did not solve the BTF explicitly, we only compare our BTF results with the ground truth in Fig. 3, where we can see that they coincide to each other well.

4.2. Set 2: i-LIDS Dataset

We further use the video clips in a multiple camera tracking scenario from the international benchmark, i-LIDS dataset [13]. Five cameras cover the airport area, and the estimated transition time distribution from two views of them are shown in Fig. 4. We select a video clip which contains 84 people with 30% unmatched outliers. The ground truth has multiple modes since people tend to travel with varying speeds due to different amounts of the carried luggage. Fig. 4 demonstrates that our method outperforms all the other approaches. The BTF estimation also achieves a good approximation of the ground truth (see Fig. 3).

TABLE I
AVERAGE ERROR

Error	Method				
	Proposed	Makris [4]	Tieu [5]	Gilbert [6]	Huang [7]
Set 1	0.011	0.24	0.2	0.38	0.066
Set 2	0.17	1.49	1.5	1.28	0.46

Table I shows the quantitative results of the above two dataset. The error is computed via the Euclidean distance between the results and the ground truth. Since the distribution of set 2 is more complicated than set 1, it has higher error value.

4.3. Error vs. The Outlier Percentage

Fig. 5 shows the errors of the estimation of the transition time distribution under different portions of the outliers based on set 1. As expected, the error increases as the percentage of the outliers in the training data raises. For simplicity, since the performance from Huang's method [7] is more reliable than the other compared methods, we only include it for comparison. Our method achieves better accuracy. For example, one can see that when the error value is around 0.1, our approach can tolerate 40% outliers instead of 25% by using the method in [7].

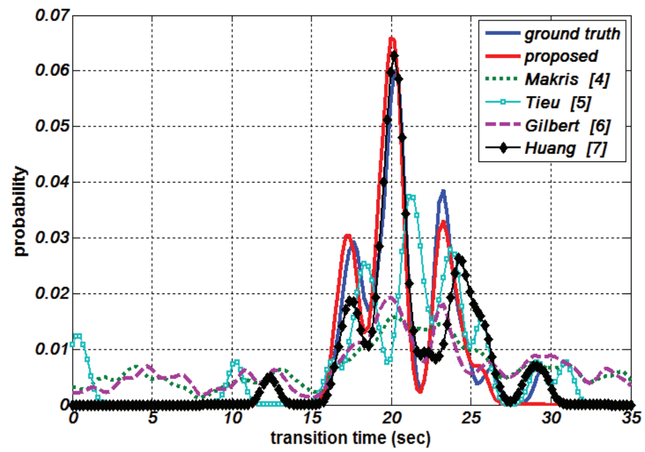

Figure 4. Transition time distribution for i-LIDS dataset.

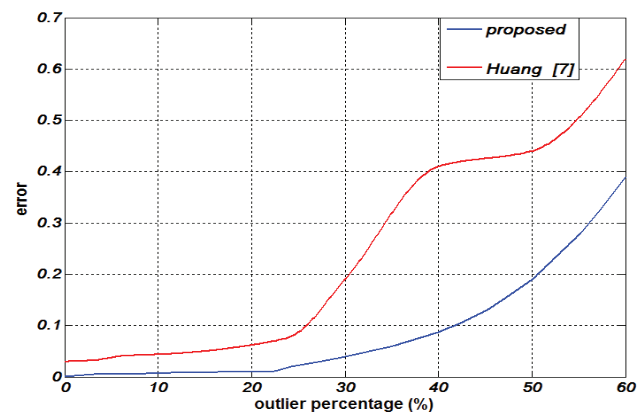

Figure 5. The error of the estimation of the transition time distribution against the outlier percentage. Blue: proposed. Red: Huang's [7]

5. CONCLUSION

In this paper, the camera link model estimation is formulated as an optimization problem and the deterministic annealing combined with the barrier method is employed to effectively find the optimal solution. The time and color information are both considered in the objective function to extract the transition time distribution and BTF simultaneously. The promising experimental results and comparative studies demonstrate that our approach is able to generate dependable model even in presence of the outliers. In the future, we will utilize more features and try to include other transfer functions in the camera link model. Finally, this model will be applied to a multiple cameras tracking system.

6. REFERENCES

[1] K. Chen, C. Lai, Y. Hung, and C. Chen, "An adaptive learning method for target tracking across multiple cameras," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition,* pp. 1-8, 2008.
[2] O. Javed, K. Shafique, Z. Rasheed, and M. Shah, "Modeling inter-camera space-time and appearance relationships for tracking across non-overlapping views," *Computer Vision and Image Understanding,* pp. 146-162, 2008.
[3] T. D'Orazio, P. Mazzeo, and P. Spagnolo, "Color brightness transfer function evaluation for non overlapping multi camera tracking," *ACM/IEEE Intl. Conf. on Distributed Smart Cameras,* pp. 1-6, 2009.
[4] D. Makris, T. Ellis, and J. Black, "Bridging the gaps between cameras," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 2, pp. 205-210, 2004.
[5] K. Tieu, G. Dalley, and W. Grimson, "Inference of non-overlapping camera network topology by measuring statistical dependence," *Proc. IEEE Conf. on Computer Vision*, vol. 2, pp. 1842-1849, 2004.
[6] A. Gilbert and R. Bowden, "Tracking objects across cameras by incrementally learning inter-camera colour calibration and patterns of activity," *Proc. ECCV*, pp. 125-136, 2006.
[7] C-C. Huang, W-C. Chiu, S-J. Wang, and J-H. Chang, "Probabilistic modeling of dynamic traffic flow across non-overlapping camera views," *IEEE Intl. Conf. on Pattern Recognition,* pp. 3332-3335, 2010.
[8] J. Maciel and J. Costeira, "A global solution to sparse correspondence problems," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 187-199, Feb 2003.
[9] H. Chiu and A. Rangarajan, "A new point matching algorithm for non-rigid registration," *Computer Vision and Image Understanding,* vol. 89, pp. 114-141, 2003.
[10] K. Rose, "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems," *Proceedings of the IEEE,* vol. 86, no. 11, pp. 2210-2239, 1998.
[11] S. Boyd and L. Vandenberghe, "Ch11: The barrier method." in Convex Optimization," *Cambridge University Press,* 2004.
[12] C. Kuo, C. Huang, and R. Nevatia, "Inter-camera association of multi-target tracks by on-line learned appearance affinity models," *Proc. ECCV*, 2010.
[13] UK Home Office, "The Image Library for Intelligent Detection Systems (i-LIDS): Multiple Camera Tracking (MCT)," 2008. *http://www.homeoffice.gov.uk/science-research/hosdb/i-lids/*