Solving Person Re-identification in Non-overlapping Camera using Efficient Gibbs Sampling

Vijay John
v.c.k.john@uva.nl

Gwenn Englebienne
g.englebienne@uva.nl

Ben Krose
b.j.a.krose@uva.nl

University of Amsterdam
Amsterdam, Netherlands

Person re-identification, or inter-camera data association, is the task of identifying people in different camera views in a network. With the recent interest in security and surveillance, camera networks are widely deployed with non-overlapping field of views (FOV). Typically, appearance-based methods are used for person re-identification using visual information corresponding to people extracted from the video sequences [1]. Appearance-based methods, though widely used, exhibit several challenging issues including spatio-temporal appearance variations, changing lighting conditions across different camera views, camera response variations and high computational complexity. The high computational complexity arises from the need to compare a unknown query person, or observation, with all other observations in the network [3]. In literature, researchers have sought to address the issues with illumination and camera response variation by using illumination invariant feature descriptors [1], by modeling the inter-camera variations in the network [2] or by colour calibration algorithms [4].

In our work, belonging to the second literature class, we model the appearance with the illumination variation and camera gain in the network. Unlike the works discussed so far [2], which model the inter-camera variations, we model the person's appearance in terms of the absolute illumination and gain associated with each camera in the network using a graphical model. To account for the high computational complexity, a novel constant time Gibbs sampling framework is proposed to perform person re-identification, and subsequently the person's appearance and camera properties are learnt in closed form. Our main contribution to literature are the following: modeling the absolute illumination variation and camera gain in the network; efficient constant time Gibbs sampling reducing the computational complexity.

Our problem is formulated as follows: given a set of observations $\mathcal{X} = \{\mathbf{x}_i\}_{i=1}^{N}$ across multiple camera views, where each observation corresponds to a person's complete trajectory within a camera's FOV, person re-identification can be defined as the problem of identifying the set of indicator variables $\mathbf{z} = \{z_i\}_{i=1}^{N}$ identifying the person associated with each observation in $\mathcal{X}$. To identify the label $z_i \in [1,\ldots,Z]$ associated with each trajectory, we use a combination of appearance features and transition probabilities between the cameras within a Bayesian inference framework. In our probablistic graphical model (Fig. 1), each observation $\mathbf{x}_i = \{l_i, e_i, t_i, \mathbf{a}_i\}$ consists of: $l_i \in [1,\ldots,L]$ the camera that records the observation; the time of entry e_i in a camera's FOV; the time of leaving the camera's FOV t_i; and the observed appearance features $\mathbf{a}_i$ corresponding to the raw RGB colour values within the bounding box of the detected person averaged over the entire trajectory. We model the appearance of a person in a camera's view as a function of the camera's properties and the person's "absolute" appearance, and we model the transition of a person from one camera to the next in terms of transition probabilities between cameras and the time it takes to move from one camera to another. The likelihood $p(\{\mathbf{x}_i\}_{i=1}^{N}|\{z_i\}_{i=1}^{N})$ is defined as

$$= \prod_{j=1}^{N} p(l_j|\{l_i\}_{i=1}^{j-1}, \{z_i\}_{i=1}^{j})\, p(e_j|\{t_i\}_{i=1}^{j-1}, \{z_i\}_{i=1}^{j})\, p(\mathbf{a}_j|z_j, l_j)\,. \quad (1)$$

Here, $p(\mathbf{a}_j|z_j, l_j)$ is modelled as $\mathbf{a}_j = g_{l_j}(\mathbf{r}_{z_j} + \mathbf{w}_{l_j})$, where g_l is the multiplicative gain constant of camera l, $\mathbf{r}_z$ is the RGB color model, averaged over the entire trajectory, $\mathbf{w}_l$ is the illumination noise associated with camera l. It is clear from its structure that this model does not allow for efficient inference, since the Markov blanket of any observation is the complete set of observations and indicator variables preceding it. Yet if the latent indicator variables are known, the observations of a person become independent of all other persons, and the model becomes much simpler (see Fig. 1(b)). We therefore use Gibbs sampling to estimate $\mathbf{z}$,

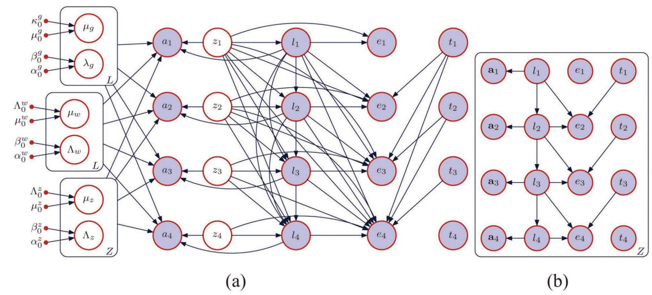

Figure 1: (a) Full graphical model of our probabilistic person re-identification algorithm and (b) Graphical model if the latent variables are known.

where each dimension of the sample $\mathbf{z}$ is sampled in alternation, according to the proposal distribution $p(z_i|\mathbf{z}_{\neg i}, \mathcal{X})$ we use $\mathbf{z}_{\neg i}$ to denote $\mathbf{z} \setminus z_i$, the set of all labels except label i. This proposal distribution leads to accepting samples with a probability of one, thereby leading to a very efficient sampling mechanism. Typically, the proposal distribution is computed as follows.

$$p(z_i|\mathbf{z}_{\neg i}, \mathcal{X}) = \frac{p(\mathcal{X}|\mathbf{z})\; p(\mathbf{z})}{\sum_{z_i=1}^{N} p(\mathcal{X}|\mathbf{z})\; p(\mathbf{z})}, \quad (2)$$

Generally, $p(\mathcal{X}|\mathbf{z})$ can be computed in linear time of the number of observations, and the probability $p(z_i|\mathbf{z}_{\neg i}, \mathcal{X})$ can also be computed in linear time. This leads to a scheme where the cost of obtaining a new sample is quadratic in the number of observations, since we need to look at all dimensions of $\mathbf{z}$. In our algorithm, we compute the prior probability over the object associations in constant time. Consequently, the conditional probability $p(z_i|\mathcal{X}, \mathbf{z}_{\neg i})$ can also be computed in constant time with a little bookkeeping. Thus reducing the computational complexity.

For a given sample of latent indicator variables, efficient Bayesian inference can be performed in our graphical model. More specifically, we obtain closed form solutions for the camera gain, illumination variation and the appearance parameters which are then used to perform analytical updates in our graphical model.

We evaluated our method on two real-world datasets with multiple non-overlapping ceiling mounted cameras and compared our performance with the algorithm proposed by Pasula et al. [3]. Additionally, we evaluate the importance modeling the camera parameters by comparing it to direct modeling of the appearance. We show that our algorithm demonstrates better re-identification accuracy and computational complexity than [3]. We also show that our proposed appearance model with gain and illumination components performs better than the modified appearance model without the gain and illumination components.

[1] Federica Battisti, Marco Carli, Giovanna Farinella, and Alessandro Neri. Target re-identification in low-quality camera networks. In *Image Processing: Algorithms and Systems*, 2013.

[2] Andrew Gilbert and Richard Bowden. Tracking objects across cameras by incrementally learning inter-camera colour calibration and patterns of activity. In *ECCV*, 2006.

[3] Hanna Pasula, Stuart Russell, Michael Ostland, and Ya'acov Ritov. Tracking many objects with many sensors. In *Int. Joint Conf. on Artificial Intelligence (IJCAI)*, pages 1160–1171, 1999.

[4] Fatih Murat Porikli. Inter-camera color calibration by correlation model function. In *ICIP*, 2003.