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

Abstract

This paper proposes a novel probabilistic approach for appearance-based person re-identification in non-overlapping camera networks. It accounts for varying illumination, varying camera gain and has low computational complexity. More specifically, we present a graphical model where we model the person's appearance in addition to camera illumination and gain. We analytically derive the solutions for the person's appearance and camera properties, and use a novel constant time Gibbs sampling scheme to estimate the identification labels. We validate our algorithm on two indoor datasets and perform a comparative analysis with existing algorithms. We demonstrate significantly increased re-identification accuracy in addition to significantly reducing the computational complexity on our datasets.

1 Introduction

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). Consequently, person re-identification in non-overlapping camera networks has gained prominent attention in the vision community. Typically, visual information corresponding to people are extracted from the video sequences as features or signatures for person re-identification. Such visual information based methods are referred to as appearance-based methods and are either single shot, designed for single video frames, or multiple shot methods, designed for video sequences [12][4]. 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. In the work by Pasula et al. [12], the Metropolis-Hastings algorithm is used to approximate the distribution of person identity labels, with the appearance model being marginalised and observations

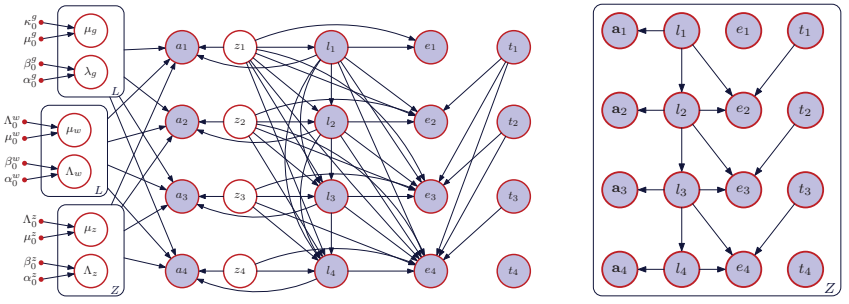
considered to be dependent. Consequently, the computation of the conditional posterior distribution is shown to scale super-exponentially with the number of observations [11].

In this paper, we propose a novel appearance-based multiple shot person re-identification algorithm that addresses the issues of illumination variation, camera gain variation, and high computational complexity in non-overlapping camera networks. In literature, researchers have sought to address the issues with illumination and camera response variation by using illumination invariant feature descriptors [2, 3, 8], by modeling the inter-camera variations in the network [1, 7, 9, 10, 14, 15] or by colour calibration algorithms [13]. In our work, belonging to the second literature class, we model the appearance with the illumination variation and camera gain in the network. Consequently, we highlight the key literature involved in modeling the network variations. Firstly, in the work by Javed et al. [10], the authors learn a inter-camera brightness transfer function in a low-dimensional subspace to account for the illumination variation. More specifically, probabilistic PCA is used to learn the subspace of Brightness Transfer Functions (BTF) for a set of known training pairs. To perform re-identification, the BTF for the candidate pair is learnt and mapped to the subspace for matching. While reporting good accuracy, the main drawback of their approach is the need for a large training dataset, which each person being sufficiently represented with diverse brightness values [14]. This issue is addressed by Prosser et al. [14], where the authors propose a modified BTF, the cumulative BTF (CBTF), by accumulating the brightness value of the entire training dataset, before learning the BTF. Consequently, they report better model estimation with comparatively sparser training dataset, without the full range of brightness values. However, the authors employ a greedy search-based re-identification scheme resulting in very high computational complexity. A similar high computational complexity is reported in the work by Gilbert et al. [7], where the inter-camera variations are modeled using an adaptive transformation matrix with a very large training dataset.

Unlike the works discussed so far [7, 10, 14, 15], 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. 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, unlike the inter-camera network variations; efficient constant time Gibbs sampling reducing the computational complexity. The structure of the paper is as follows, we present our algorithm in Section 3. In Section 4, we discuss our experimental results, before presenting our conclusion with direction to future work in Section 5.

2 The Model

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 $z_i \in [1,\dots,Z]$ associated with each trajectory, we use a combination of appearance features and transition probabilities between the cameras. To address the issues associated with appearance-based methods (Section 1) in our proposed person re-identification algorithm, we model each person's appearance using camera-specific illumination and camera gain. We identify the indicator labels by performing Bayesian inference using Gibbs sampling.



(a) (b)

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

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. For the purpose of this paper, the people's appearance and the cameras' properties are learnt online, while the transitions between cameras are a function of the environment and are provided *a priori*. Additionally, the number of people (Z) are also provided *a priori*. The corresponding graphical model is depicted in Fig. 1. The likelihood is defined as

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

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-based appearance model, averaged over the entire trajectory, $\mathbf{w}_l$ is the illumination noise associated with camera l, and the terms are distributed as:

$$g_l \sim \text{Gamma}(\alpha_l^g, \beta_l^g), \text{ which we approximate as } \mathcal{N}(\mu_l^g, (\Lambda_l^g)^{-1}) \quad (2)$$

$$\mathbf{r}_z \sim \mathcal{N}(\mu_z, (\Lambda_z)^{-1}) \quad (3)$$

$$\mathbf{w}_l \sim \mathcal{N}(\mu_l^w, (\Lambda_l^w)^{-1}) \quad (4)$$

The transitions between cameras are modelled as

$$l_j \mid \{l_i\}_{i=1}^{j-1}, \{z_i\}_{i=1}^{j-1} \sim \text{Mult}(l_j; \theta_{l_i}), i : z_i = z_j \wedge z_k \neq z_j, i < k < j \quad (5)$$

$$e_j \mid \{t_i\}_{i=1}^{j-1}, \{z_i\}_{i=1}^{j} \sim \mathcal{N}(e_j - t_i; \mu_{l_i,l_j}, \Lambda_{l_i,l_j}^{-1}), i : z_i = z_j \wedge z_k \neq z_j, i < k < j \quad (6)$$

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)). Note that the conditional dependencies in Fig. 1(b) are valid when the corresponding observations have the same person label. We therefore use sampling to estimate $\mathbf{z}$.

3 Gibbs Sampling-based Person Re-identification

Gibbs sampling [6] is a form of MCMC sampling 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.

$$p(z_i|\mathbf{z}_{\neg i}, \mathcal{X}) = \frac{p(\mathcal{X}|\mathbf{z})\ p(\mathbf{z})}{\sum_{z_i=1}^{N} p(\mathcal{X}|\mathbf{z})\ p(\mathbf{z})}, \tag{7}$$

, where $p(\mathcal{X}|\mathbf{z})$ and $p(\mathbf{z})$ 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 $\mathbf{z}$. However, if the prior probability over the object associations can be computed in constant time, the conditional probability $p(z_i|\mathcal{X}, \mathbf{z}_{\neg i})$ can also be computed in constant time with a little bookkeeping.

3.1 Constant time Gibbs Sampling

The faster sampling mechanism works as follows. Let $\mathbf{z}$ denote the set of persons associated with the observations $\mathcal{X}$, and b_i be the index of the previous observation associated with the same person z_i. Similarly, let f_i indicate the next observation associated with that person. That is,

$$b_i \triangleq \arg\max_{j<i} (z_j = z_i) \text{ and} \tag{8}$$

$$f_i \triangleq \arg\min_{j>i} (z_j = z_i). \tag{9}$$

If we know $p(\mathcal{X}|\mathbf{z})$, we can compute the probability of the observations given a different label z_i' for observation i, $p(\mathcal{X}|\mathbf{z}_{\neg i}, z_i')$, in constant time as follows:

$$p(\mathcal{X}|\mathbf{z}_{\neg}, z_i') = p(\mathcal{X}|\mathbf{z}) \times \frac{p(\mathbf{x}_i|z_i', \mathbf{x}_{b_i'}) p(\mathbf{x}_{f_i}|z_{f_i}, \mathbf{x}_{b_i}) p(\mathbf{x}_{f_i'}|z_i', \mathbf{x}_i)}{p(\mathbf{x}_i|z_i, \mathbf{x}_{b_i}) p(\mathbf{x}_{f_i}|z_{f_i}, \mathbf{x}_i) p(\mathbf{x}_{f_i'}|z_i', \mathbf{x}_{b_{f_i'}})}. \tag{10}$$

Here, we take the probability of the sequence of observations given the associated labels, divide out the terms that were affected by z_i and multiply in the terms that are affected by z_i'. In practice, of course, one would work with log-probabilities, so that the division operations would become subtractions and the multiplications become additions. Sampling from the conditional probability distribution $p(z_i|\mathcal{X}, \mathbf{z}_{\neg i})$ is an $O(Z)$ operation, constant in N. Each sample therefore consists of the sequence of labels, $\mathbf{z}$, the sequence of "back pointers", $\{b_i\}_{i=1}^N$, and the sequence of "forward pointers", $\{f_i\}_{i=1}^N$. When we sample the value of the

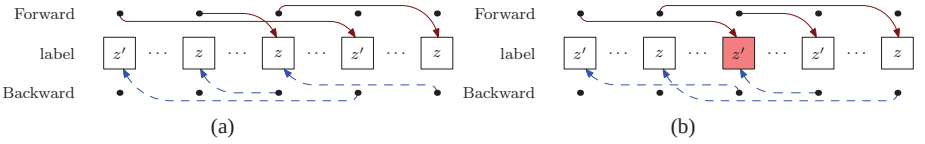


Figure 2: Illustration of the bookkeeping used for constant-time sampling. For clarity, only the modified bookkeeping variables are shown; all other variables remain unchanged

label for observation i, we first compute the quantities that need be divided out and then modify forward and backward pointers, in order, as follows:

$$f_{b_i} \leftarrow f_i \qquad b_{f_i} \leftarrow b_i \qquad f_i \leftarrow f'_i \qquad b_i \leftarrow b_{f_i} \qquad b_{f_i} \leftarrow i \qquad f_{b_i} \leftarrow i \tag{11}$$

This is illustrated in Fig. 2, and requires one computation: for the next occurrence of the new label f'_i. We do this in constant time by maintaining a list of the next occurrence of any label at time i, but in practice it could also be implemented as a linear search, since the next occurrence of a label will typically be much closer than the last observation. Notice that extra care must be taken for the first and last observations associated with an object, where the backward and forward pointer, respectively, cannot point to anything meaningful. Sampling the label of an observation is constant in the number of observations, and linear in the number of objects. The time complexity of updating $\mathbf{z}$ is therefore $O(NZ)$.

*Initialisation.*MCMC methods require a "burn-in" period, in order to converge to the equilibrium distribution of the chain. During this period, we sample from the chain and discard the samples, which is time-consuming. How long we need to sample is hard to assess and depends both on the convergence rate of the chain and on the distribution of the initial samples. If the initial sample is close to the high-probability region of the space, less sampling will be required for the chain to converge to the desired distribution. Although we cannot sample from the "smoothed" distribution for the labels, $p(\mathbf{z}|\mathcal{X})$, we can however easily sample from the "forward" distribution, $p(z_j|\{z_i\}_{i=1}^{j-1},\mathcal{X})$, by similarly summing over all possible values of z_j. The forward distribution does not take future observations into account, but it is typically close enough to the smoothed distribution that a few tens of sampling iterations are sufficient to converge to the smoothing distribution.

3.2 Gain Conditional Distribution

For a given sample of latent indicator variables, efficient Bayesian inference can be performed in our graphical model. Firstly, using the Markov blanket in the graphical model in Fig. 1, the conditional distribution over the gain parameters for each independent camera l in the network is defined in terms of the likelihood and prior distribution in Eq. 12,

$$p(\mu_l^g,\lambda_l^g|A_l,\mu_l^w,\Lambda_l^w,\mu_z,\Lambda_z,l,\mathbf{z}) \propto p(A_l|\mu_l^g,\lambda_l^g,\mu_l^w,\Lambda_l^w,\mu_z,\Lambda_z,\mathbf{z},l)\; p(\mu_l^g,\lambda_l^g)\;, \tag{12}$$

where $A_l = \{a_i\}_{i:l_i=l}$ represents the set of appearance observations seen in camera l. The likelihood distribution in Eq. 12 is formulated as a Gaussian distribution given as,

$$p(A_l|\mu_l^g,\lambda_l^g,\mu_l^w,\Lambda_l^w,\mu_z,\Lambda_z,\mathbf{z},l) = \prod_{i:l_i=l} N(a_i|\mu_l^g(v_l^z),\lambda_l^g(\varepsilon_l^z)) \tag{13}$$

, where $v_l^z = \mu_z + \mu_l^w$ and $\varepsilon_l^z = \Lambda_z + \Lambda_l^w$. Given the Gaussian likelihood distribution with unknown mean and unknown precision, the normal-gamma distribution is chosen as the conjugate prior distribution (Eq. 14).

$$p(\mu_l^g, \lambda_l^g | \alpha_0^g, \kappa_0^g, \beta_0^g, \mu_0^g) \propto \lambda_l^{g(1/2)} \exp(-\kappa_0^g \lambda_l^g / 2(\mu_l^g - \mu_0^g)^2) \lambda_l^{g(\alpha_0^g - 1)} e^{-\lambda_l^g \beta_0^g} \tag{14}$$

where, $\alpha_0^g, \kappa_0^g, \beta_0^g, \mu_0^g$ represent the prior hyper-parameters. Since the posterior distribution for a conjugate prior share the same functional form, $\alpha_n^g, \kappa_n^g, \beta_n^g, \mu_n^g$ is used to represent the posterior parameters. To solve for the conditional posterior distribution, the likelihood distribution is factorised according to each person's appearance and then multiplied with the prior distribution (Eq. 14) yielding the posterior distribution. Next, we algebraically complete the square with respect to μ_l^g, as described in [5], and derive the analytical solutions for the posterior parameters, given as,

$$\mu_n^g = \frac{\sum_z n_l^z \overline{a}^T (\lambda_l^g(\varepsilon_l^z)) \mu_z + \sum_z n_l^z \overline{a}^T (\lambda_l^g(\varepsilon_l^z)) \mu_l^w + \kappa_0^g \lambda_l^g \mu_0^g}{\sum_z n_l^z (\mu_z^T (\lambda_l^g(\varepsilon_l^z)) \mu_z + \mu_l^{wT} (\lambda_l^g(\varepsilon_l^z)) \mu_l^w + 2\mu_z^T (\lambda_l^g(\varepsilon_l^z)) \mu_l^w) + \kappa_0^g \lambda_l^g} \tag{15}$$

$$\kappa_n^g = \sum_z n_l^z (\mu_z^T (\lambda_l^g(\varepsilon_l^z)) \mu_z + \mu_l^{wT} (\varepsilon_l^z) \mu_l^w + 2\mu_z^T (\varepsilon_l^z) \mu_l^w) + \kappa_0^g \tag{16}$$

$$\beta_n^g = \beta_0^g + \frac{\sum_z S_z^2 (n_l^z - 1)(\varepsilon_l^z)}{2} + \frac{\kappa_0^g \mu_0^{g2} \left(\sum_z n_l^z (v_l^z)^T (\varepsilon_l^z)(v_l^z)\right) \ldots}{2(\sum_z n_l^z (\mu_z^T (\varepsilon_l^z) \mu_z + \mu_l^{wT} (\varepsilon_l^z) \mu_l^w + 2\mu_z^T (\varepsilon_l^z) \mu_l^w) + \kappa_0^g)} \tag{17}$$

$$\frac{+\kappa_0^g \left(\sum_z n_l^x z \overline{a}^T (\varepsilon_l^z) \overline{a} - 2\mu_0^g \sum_z n_l^z \overline{a}^T (\varepsilon_l^z) \mu_l^w - 2\mu_0^g \sum_z n_l^z \overline{a}^T (\varepsilon_l^z) \mu_z\right)}{2(\sum_z n_l^z (\mu_z^T (\varepsilon_l^z) \mu_z + \mu_l^{wT} (\varepsilon_l^z) \mu_l^w + 2\mu_z^T (\varepsilon_l^z) \mu_l^w) + \kappa_0^g)}$$

$$\alpha_n^g = \alpha_0^g + \frac{3N_l}{2} \tag{18}$$

where n_l^z represents the number of trajectories for each person, N_l represents the number of trajectories observed by camera l, $\overline{a} = \sum_i a_i / n$ corresponds to the empirical mean and $S^2 = (\sum_{i=1}^n (a_i - \overline{a})^T (a_i - \overline{a}))/n$ represents each person's empirical covariance matrix.

3.3 Illumination Variation Conditional Distribution

The conditional distribution over the illumination variation per camera is defined, for easier analytical derivation, by modeling the conditional distribution separately for the mean and precision components. The conditional distribution for the *illumination precision*, in terms of the likelihood and prior distributions, using the Markov blanket in Fig. 1, is given by

$$p(\Lambda_l^w | A, \mu_l^g, \lambda_l^g, \mu_l^w, \mu_z, l, z) = p(A | \mu_l^g, \lambda_l^g, \mu_l^w, \Lambda_l^w, \mu_z, \Lambda_z, z, l) p(\Lambda_l^w) \,, \tag{19}$$

where the likelihood distribution is the same as defined in Eq. 13. The conjugate prior and the subsequent posterior for unknown precision and known mean is given by the Wishart distribution, as

$$p(\Lambda_l^w | \alpha_0^w, \beta_0^w) \propto |\Lambda_l^w|^{\frac{(\alpha_0^w - 4)}{2}} \exp\left(-\frac{1}{2} \mathrm{tr}(\beta_0^w \Lambda_l^w)\right) \tag{20}$$

where α_0^w, β_0^w represent the prior hyper-parameters and α_n^w, β_n^w are used to represent the posterior parameters. To solve for the conditional distribution, we factorize the likelihood in terms of the appearance, multiply the resultant distribution with the prior and follow the algebraic derivations in [5] to derive the precision hyper-parameters, given as,

$$\alpha_n^w = \alpha_0^w + N_l + 1 \tag{21}$$

$$\begin{aligned}\beta_n^w = \beta_0^w + \sum_z (\sum_{i=1}^{n_z^l} (a_i - \overline{a})^T (\lambda_l^g)(a_i - \overline{a}) + n_l^z \overline{a}^T \lambda_l^g \overline{a} - n_l^z 2\overline{a}^T (\lambda_l^g)\mu_l^g \mu_z \\ \ldots - n_l^z 2\overline{a}^T (\lambda_l^g)\mu_l^g . \mu_l^w + n_l^z \mu_l^{g2} \mu_z^T (\lambda_l^g)\mu_z + n_l^z \mu_l^{g2} \mu_l^{wT} (\lambda_l^g)\mu_l^w + n_l^z 2\mu_l^{g2} \mu_z^T (\lambda_l^g)\mu_l^w)\end{aligned} \tag{22}$$

Illumination Mean. The conditional distribution for the *illumination mean* is modeled using the Markov blanket (Fig. 1), as

$$p(\mu_l^w | a, \mu_l^g, \lambda_l^g, \Lambda_l^w, \mu_z, l, z) = p(a | \mu_l^g, \lambda_l^g, \mu_l^w, \Lambda_l^w, \mu_z, \Lambda_z, z, l) p(\mu_l^w) \tag{23}$$

with the likelihood distribution being defined in Eq. 13. The conjugate prior and the subsequent posterior for unknown mean and known precision is given by the multivariate Gaussian distribution as

$$p(\mu_l^w | \mu_0^w, \Lambda_0^w) \propto |\Lambda_0^w|^{\frac{1}{2}} \exp\left(-\frac{1}{2} (\mu_l^w - \mu_0^w)^T \Lambda_0^w (\mu_l^w - \mu_0^w) \right) \tag{24}$$

where μ_0^w, Λ_0^w represent the prior hyper-parameters and μ_n^w, Λ_n^w is used to represent the posterior hyper-parameters. Again, following the algebra in [5], we factorise the likelihood in terms of each person's appearance, multiply with the prior, complete the squares with respect to μ_l^w and derive the analytical solution for the mean hyper-parameters, given as,

$$\mu_n^w = \frac{\sum_z n_l^z \lambda_l^g \mu_l^g \Lambda_z \overline{a} + \sum_z n_l^z \mu_l^g \lambda_l^g \Lambda_l^w \overline{a} - \sum_z n_l^z \mu_l^{g2} \lambda_l^g \Lambda_z \mu_z - \sum_z n_l^z 2\mu \lambda_l^g \Lambda_l^w \mu_z + \Lambda_0^w \mu_0^w}{\sum_z n_l^z \mu_l^{g2} \lambda_l^g \Lambda_z + \sum_z n_l^z \mu_l^{g2} \lambda_l^g \Lambda_l^w + \Lambda_0^w} \tag{25}$$

$$\Lambda_n^w = \sum_z n_l^z \mu_l^{g2} \lambda_l^g \Lambda_z + \sum_z n_l^z \mu_l^{g2} \lambda_l^g \Lambda_l^w + \Lambda_0^w \tag{26}$$

3.4 Appearance Conditional Distribution

Similar to the derivation of illumination variation, we model the conditional distribution separately for the mean and precision components using the Markov blanket in the graphical model (Fig. 1). The choice of conjugate priors for the appearance mean and precision derivations are also similar to the illumination variation derivations (multi-variate Gaussian and Wishart distribution), resulting in similar analytic derivations. Thus, we directly give the hyper-parameter solutions. First, the analytic solution for the appearance precision hyper-parameters $p(\Lambda_z | \alpha_n^z, \beta_n^z)$ are given in Eq. 27 and Eq. 28, where N_z represents the number of trajectories corresponding to person z. While the analytic solution for the appearance mean hyper-parameters $p(\mu_z | \mu_n^z, \Lambda_n^z)$ are given in Eq. 30.

$$\alpha_n^z = \alpha_0^z + N_z + 1 \tag{27}$$

$$\beta_n^z = \beta_0^z + \sum_l (\sum_{i=1}^{n_z^l} \lambda_l^g (a_i - \overline{a})^T (a_i - \overline{a}) + n_z^l \overline{a}^T \lambda_l^g \overline{a} - n_z^l 2\overline{a}^T \lambda_l^g \mu_l^g \mu_z$$
$$\ldots - n_z^l 2\overline{a}^T \lambda_l^g \mu_l^g \mu_l^w + n_z^l \mu_l^{g2} \mu_z^T \lambda_l^g \mu_z + n_z^l \mu_l^{g2} \mu_l^{wT} \lambda_l^g \mu_l^w + n_z^l 2\mu_l^{g2} \mu_z^T \lambda_l^g \mu_l^w) \quad (28)$$

$$\mu_n^z = \frac{\sum_l n_z^l \lambda_l^g \mu_l^g \Lambda_z \overline{a} + \sum_l n_z^l \mu_l^g \lambda_l^g \Lambda_l^w \overline{a} - \sum_l n_z^l \mu_l^{g2} \lambda_l^g \Lambda_z \mu_l^w - \sum_l n_z^l \mu_l^{g2} \lambda_l^g \Lambda_l^w \mu_l^w + \Lambda_0^z \mu_0^z}{\sum_l n_z^l \mu_l^{g2} \lambda_l^g \Lambda_z + \sum_l n_z^l \mu_l^{g2} \lambda_l^g \Lambda_l^w + \Lambda_0^z} \quad (29)$$

$$\Lambda_n^z = \sum_l n_z^l \mu_l^{g2} \lambda_l^g \Lambda_z + \sum_l n_z^l \mu_l^{g2} \lambda_l^g \Lambda_l^w + \Lambda_0^z \quad (30)$$

4 Experiments

We evaluated our method on two real-world datasets and compared our performance with the algorithm proposed by Pasula et al. [12]. 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 [12]. We also show that our proposed appearance model with gain and illumination components performs better than the naive appearance model. More details on the naive appearance model are provided below.

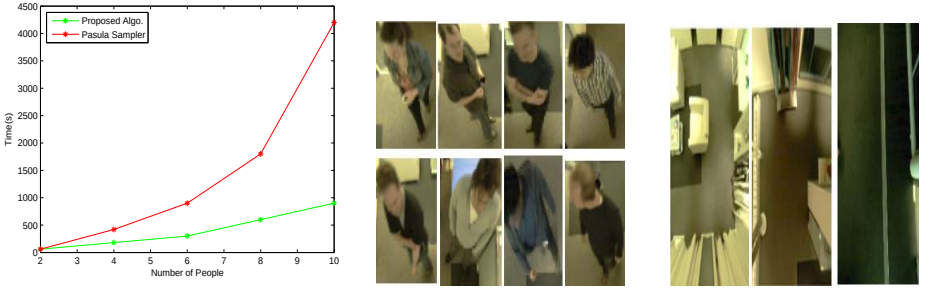
Dataset. In our experiments, we acquired two indoor studio datasets of multiple walking subjects, named, dataset-1 and dataset-2 using ceiling mounted colour depth cameras @15Hz. dataset-1 consists of two video sequences with 5 and 10 subjects acquired with 5 cameras, while dataset-2 consists of one video sequence with 5 subjects acquired with 13 cameras. We implemented our algorithm in C++ in Linux with 3.49GHz processor. Examples of subjects and the camera FOV from dataset-2 are shown in Fig. 3(b-c). The observation tracks used in our algorithm were obtained using our in-house multi-person tracking algorithm.

Comparative Result. We compare the performance of our algorithm with the algorithm proposed by Pasula et al. [12], which we re-implemented. As shown in Table 1, we can see that our proposed algorithm performs significantly better than [12] across all three test sequences. The re-identification accuracy is measured as the percentage of observations that have been assigned to the correct label. Table 1 shows the mean and standard deviation of the re-identification accuracy averaged over 5 trials, since the algorithms are stochastic in nature. To evaluate the computational complexity, we consider the sequence with 10 subjects in dataset-2 with varying number of people and measure the computational time of our proposed algorithm and [12], shown in Fig. 3. We can clearly see the significant improvement in computational complexity with our proposed algorithm. Moreover, as discussed in [11], the exponential scaling of the computational complexity of [12] is also observed.

Camera Parameters In our second experiment, we evaluate the algorithm's performance without camera model. More specifically, we directly use the raw RGB ($\mathbf{a}_i$) (Sec. 3) observations without modeling the camera-specific gain and illumination parameters, which we refer to as the naive appearance model. The algorithms are then evaluated on the test sequences and the results are shown in Table 1, where our proposed algorithm performs significantly

	Proposed Algo	Naive App Model	[12]
DataSet-1	87.5+4.3%	75+5.1%	67.5+6.1%
DataSet-2 (Seq: 5 sub)	86+5.2%	74.3+4.6%	65+8.1%
DataSet-2 (Seq: 10 sub)	84+4.1%	73.2+5.3%	62.5+7.3%

Table 1: The mean and std. dev of the re-identification accuracy over the three test sequences are shown. Please refer to the text for details about the naive appearance model.



(a) (b) (c)

Figure 3: (a) Computational Complexity with varying number of people. (b) Sample subjects from dataset-2. (c) Sample images from different cameras in dataset-2 with varying illumination.

better than the standard appearance model. Next, we evaluate our proposed algorithm with varied number of frames on the 10 subject sequence in dataset-2 and measure the corresponding accuracy. As shown in Table 2(b), we can see that our proposed algorithm demonstrates a steady increase in re-identification accuracy with increased number of frames, which can be attributed to the improved estimate of the appearance model with the increasing number of frames.

5 Conclusion

We have proposed a novel appearance-based person re-identification algorithm for a camera network addressing the challenging issues of camera gain variation, illumination variation and high computational complexity. In our proposed solution, we have modeled our appearance model incorporating the network gain and illumination variation. Additionally, we have proposed a novel sampling approach by maintaining a little bookkeeping for re-identification. We have evaluated our algorithm on two real-world datasets and demonstrate that our proposed algorithm performs better than comparative algorithms both in terms of re-identification accuracy and computational complexity. Additionally, we demonstrate the advantage of incorporating the network gain and illumination component within the appear-

Frames	200	600	1000	1400	2000
Accuracy (%)	72.5+5.6%	77+5.2%	81.6+4.6%	82.1+3.9%	84+4.1%

Table 2: Mean and std.dev of the re-identification accuracy with varying number of frames

ance model. In our future work, we would like to test our algorithm with a larger dataset and outdoor sequences. Additionally, we would like to learn the transitions between cameras in the network automatically.

References

[1] Slawomir Bak, Guillaume Charpiat, Etienne Corvee, Francois Bremond, and Monique Thonnat. Learning to match appearances by correlations in a covariance metric space. In *ECCV*, 2012.

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

[3] Etienne Corvée, Slawomir Bak, and François Brémond. People detection and re-identification for multi surveillance cameras. In *VISAPP*, 2012.

[4] Michela Farenzena, Loris Bazzani, Alessandro Perina, Vittorio Murino, and Marco Cristani. Person re-identification by symmetry-driven accumulation of local features. In *CVPR*, 2010.

[5] Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. *Bayesian Data Analysis, Second Edition*. Chapman and Hall/CRC, 2003. ISBN 158488388X.

[6] Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on PAMI*, 6(1):721–741, 1984.

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

[8] Omar Hamdoun, Fabien Moutarde, Bogdan Stanciulescu, and Bruno Steux. Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences. In *ICDSC*, 2008.

[9] Martin Hirzer, PeterM. Roth, Martin Kostinger, and Horst Bischof. Relaxed pairwise learned metric for person re-identification. In *ECCV*, 2012.

[10] Omar Javed, Khurram Shafique, and Mubarak Shah. Appearance modeling for tracking in multiple non-overlapping cameras. In *CVPR*, 2005.

[11] Hanna Pasula. Identity uncertainty. In *Doctoral Thesis, University of California, Berkley*, 2003.

[12] 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.

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

[14] B. Prosser, S. Gong, and T. Xiang. Multi-camera matching using bi-directional cumulative brightness transfer functions. In *Proc. BMVC*, pages 64.1–64.10, 2008.

[15] Bryan Prosser, Shaogang Gong, and Tao Xiang. Multi-camera matching under illumination change over time. In *ECCV*, 2008.