Multi-camera Matching under Illumination Change Over Time

Bryan Prosser, Shaogang Gong, and Tao Xiang

Department of Computer Science
Queen Mary, University of London
{bryan, sgg, txiang}@dcs.qmul.ac.uk

Abstract. Illumination differences between disjoint cameras can have a dramatic effect on the appearance of objects, thus increasing the difficulty of multi-camera object association. Although methods to model these inter-camera illumination conditions exist, they often rely on static illumination conditions and are unable to copewith unpredictable illumination changes over time. In this paper we propose a novel method for multi-camera object association based on adapting a learned inter-camera illumination mapping function to new illumination conditions over time without the need for a manual training stage using new foreground objects. Comparative experiments are carried out using challenging data taken from a disjoint camera network. The results demonstrate that the proposed method outperforms a number of existing methods given changing illumination conditions.

1 Introduction

A typical video surveillance system employs a number of networked cameras, many of which have disjoint views. One of the key problems of behaviour monitoring using a networked cameras is to track people across camera views, known as the person re-identification problem. Specifically, to re-establish a match of the same person over different camera views located at different physical sites, one aims to track individuals either retrospectively or on the fly when they move through different sites. Due to considerable changes in object orientation, pose and lighting conditions between camera views, this task is non-trivial. Since real world camera networks rarely have overlapping views, the key challenge for re-identification is to establish object correspondence given these changes. Among these conditions that vary across cameras, dealing with the lighting condition change is particularly challenging. This is because light conditions at different camera views can change over time in a unknown manner. While methods exist to address the problem of illumination change between camera views [1–3], none of them consider the lighting condition changes over time. These changes at each camera view result in additional changes across cameras not modelled by existing techniques. (see examples in Figure 1). Our aim is therefore to enhance multi-camera re-identification by modelling temporal illumination changes.

Simple appearance based methods currently exist to handle the lighting condition changes between cameras. Cheng et al [3] cluster colours into a subset of *major colours*, and to alleviate the effect of illumination changes, they employ a

(a) Day 1 (b) Day 2

Fig. 1. Illumination condition can change over time especially when outdoor lighting plays a part. In this case Day 1 was a dull cloudy day and Day 2 was sunny.

histogram equalisation technique. A more sophisticated method is to model the illumination differences between each pair of camera views explicitly. Javed et al [1] proposed a subspace based colour brightness transfer function (BTF). They use probabilistic PCA to calculate the subspace of BTFs for a set of known correspondences and compare the BTF of a test object pair against this subspace to determine a correspondence. Prosser et al [4] also use a BTF-based approach but accumulate training data before computing the BTF. This Cumulative BTF (CBTF) enables sparse colour information to be preserved through the BTF calculation process. Gilbert and Bowden [5] model inter-camera colour transformations using an incrementally updated transformation matrix. A similar model was proposed in [2] without incremental learning. Instead, it employs a hardware calibration phase to learn colour differences, which is impractical because many real world systems do not offer the access to the hardware parameters.

The main shortcoming of the previous approaches is that they are trained for a single constant lighting condition at each individual view. If the illumination at any one of the cameras changes over a period of time, the BTF based approaches would require manual selection of a substantial set of corresponding object observations in each camera to re-learn the brightness mapping under the new conditions. It is therefore tedious, time-consuming, and above all is not even possible when illumination condition changes are non-gradual. The histogram equalisation technique used by Cheng et al [3] assumes some arbitrary a priori knowledge of a suitable colour mapping function which again would need to be updated manually.Gilbert and Bowden's incremental colour mapping method [5] has the potential to cope with temporal illumination changes provided such changes are gradual. However, their method requires thousands of object appearances to learn an accurate brightness mapping between camera views, which are unlikely to be available especially when the lighting conditions changes are non-gradual.

In this paper we propose a novel method for multi-camera people association based on adapting cumulative brightness transfer function (CBTF) to new illumination conditions without the need for a manual training stage using new foreground objects. This method therefore can run in real time even given con-

stantly changing lighting conditions. More specifically, by modelling the temporal changes in background illumination, the updated CBTF is estimated using a combination of the original CBTF and the background illumination changes at each camera. The proposed method is evaluated using challenging datasets obtained from a real world CCTV camera network. The results demonstrate that the adaptive CBTF estimation is accurate and the proposed method significantly outperforms existing approaches.

2 Inferring Illumination Changes Over Time

Given a pair of camera views i and j and the CBTF cf_{ij} obtained using a set of object correspondences, we aim to adaptively update the CBTF to any change of illumination condition without collecting new object correspondences. The new camera views under a different illumination condition is denoted as i' and j', and the updated CBTF $cf_{i'j'}$. This is achieved through calculating a colour mapping function for each of the two camera views over time, denoted as $f_{i'i}$ and $f_{j'j}$ respectively. This enables us to convert object images under a different illumination condition back to the illumination conditions under which the original CBTF was learned. The approach is illustrated in Figure 2.

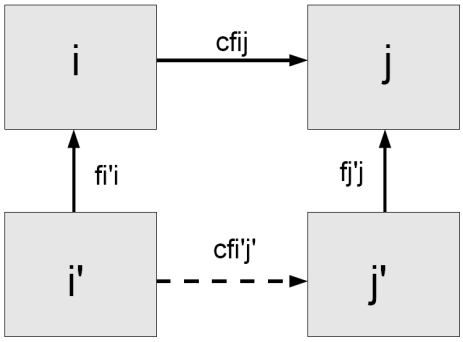


Fig. 2. Camera views i,j and i',j' under different illumination conditions. By modelling the illumination change for each camera view ($f_{i'i}$ and $f_{j'j}$) we are able to utilise the original inter-camera cf_{ij} to infer the new inter-camera $cf_{i'j'}$ without re-training.

2.1 Brightness Transfer Function

First let us formally define the multi-camera person re-identification problem. A camera network has m cameras $C_1, ..., C_m$ all of which are assumed to have non-overlapping views. We break down each view into entry/exit regions as illumination varies between both inter- and intra-camera regions. Specifically, for each of the camera views we define its set of n entry/exit regions as $E^1_{C_1}, ..., E^n_{C_1}$. We then simplify this by describing the global set of g entry/exit regions as $E_1, ..., E_g$. Next we define a set of k object observations in each entry/exit region E_i as $\{O_{i,1}, ..., O_{i,k}\}$. Using an existing single camera view tracking system we can obtain these observations by taking a colour histogram of a target object as the centre of its bounding box passes through an entry/exit region. In order to solve the multi-camera re-identification problem we form a set of correspondence

hypotheses Q where each $Q_{i,a}^{j,b}$ indicates a potential match between observations $O_{i,a}$ and $O_{j,b}$. We consider the solution space S as a set of all possible correspondence hypotheses between E_i and E_j. Assuming that an object in E_i is seen no more than once in E_j we aim to find the subset of S, s where each $Q_{i,a}^{j,b} \in s$, if and only if observations $O_{i,a}$ and $O_{j,b}$ are the same person. The solution of the multi-camera re-identification problem is then defined as $s \in S$ which maximises an observation similarity measure:

$$s = \arg\max_{s} \prod_{Q_{i,a}^{j,b} \in s} Similarity(O_{i,a}, O_{j,b}) \tag{1}$$

where $Similarity()$ is the similarity between $O_{j,b}$ and $O_{i,a}$ in the testing data.

A vital part of the re-identification process lies in the BTF. Javed et al [1] suggested that a BTF $f_{ij}(\,)$ between cameras C_i and C_j can be constructed by sampling values from a set of fixed increasing brightness levels $B_i(1), ..., B_i(d)$, and $(B_j(1), ..., B_j(d)) = (f_{ij}(B_i(1)), ..., (f_{ij}(B_i(d)))$. In the case of a common 8-bit per channel image, d is set to 256. To establish such a mapping function between views, a pair of known correspondence must be available. These correspondences are represented as normalised histograms of RGB brightness values. Computing a mapping function can be achieved as follows. It is assumed that the percentage of pixels in an observation O_i with the brightness value less than B_i is equal to the percentage of image points seen in O_j of brightness less than or equal to B_j. H_i and H_j are then defined as cumulative histograms. More specifically, for H_i each bin of brightness value $B_1, ..., B_m, ..., B_{256}$ in one of the three colour channels is obtained from the colour histogram h_i as follows:

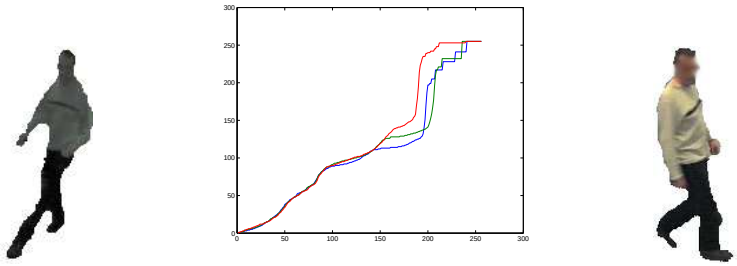
$$H_i(B_m) = \sum_{k=1}^{m} I_i(B_k) \tag{2}$$

where $I_i(B_k)$ is the count of brightness value B_k in O_i. Each bin is then normalised using the total number of pixels in O_i. $H_i(B_i)$ represents the proportion of H_i less than or equal to B_i, then $H_i(B_i) = H_j(B_j) = H_j(f_{ij}(B_i))$ and the BTF mapping function can be defined as:

$$f_{ij}(B_i) = H_j^{-1}(H_i(B_i)) \tag{3}$$

with H^{-1} representing the inverted cumulative histogram. An example BTF and the corresponding observation images can be seen in Figure 3.

In order to produce a more accurate transfer function, multiple BTFs can be estimated. Prosser et al [4] show that mean BTF based methods rely on having a consistent set of coloured individuals to accurately model the BTF and that taking the mean of a set of BTFs actually removes vital colour information that may only be contained in a small subset of the training data. Rather than computing a BTF for each training pair they accumulate the brightness values of the whole training set before the BTF computation. The cumulative histogram



(a) (b) (c)

Fig. 3. (a) and (c) show a corresponding person in 2 camera views, (b) shows the BTF converting between the two illumination conditions.

$\hat{H}_i$ of N training samples in camera view i can be computed from the brightness values $B_1, ..., B_m, ..., B_{256}$ as:

$$\hat{H}_i(B_m) = \sum_{k=1}^{m} \sum_{L=1}^{N} I_L(B_k). \tag{4}$$

After obtaining this cumulative histogram using all the training image pairs, the CBTF is computed as follows

$$cf_{ij}(B_i) = \hat{H}_j{}^{-1}(\hat{H}_i(B_i)) \tag{5}$$

This CBTF is used as the cf_{ij} from Figure 2 in the following section.

2.2 Modelling Temporal Illumination Changes

The first stage in modelling the illumination change is to derive a single median background image from each of the two datasets collected under different illumination conditions. To do this we collect a set of background images, $I_1, ..., I_n$, by using a background/foreground subtraction technique to find images containing minimal foreground regions. In our case we chose a set size of 20 images. From this set of images we compute the median RGB values at each pixel location and produce a median background image. The two median background images for the two different illumination conditions are denoted as $M_i(x, y)$ and $M_{i'}(x, y)$.

In each of the median background images we extract regions of interest R that corresponds to entry/exit regions of a camera. In our work these regions are manually defined, however we are aware that work exists to extract these automatically, such as [6, 7]. As the background of a scene may change over time due to reasons other than illumination change, e.g. the movement of a static object, we perform frame differencing to remove these areas so that they do not pollute the final colour mapping. Let $\hat{M}_i(x, y)$ denote a region of interest R after frame differencing, we have:

$$\forall x, y \in R, \hat{M}_i(x, y) = \begin{cases} M_i(x, y) \text{ if } abs(M_i(x, y) - M_{i'}(x, y)) < \sigma \\ 0 \qquad\qquad \text{otherwise} \end{cases} \tag{6}$$

where σ was typically between 30 and 50. An example of regions of interests extracted from median background images of an entry/exit region of a camera is shown in Figure 4.

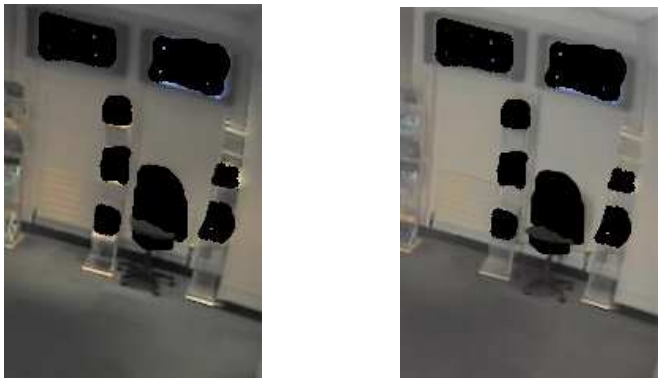


Fig. 4. Corresponding regions of interest from the same entry/exit region of a camera site on Day 1 and 2 respectively with pixels with large value changes removed. Those removed pixels correspond to an LCD display, a chair, and some magazines, all of which have been changed/moved over the two days.

From $\hat{M}_i(x, y)$ and similarly calculated $\hat{M}_{i'}(x, y)$ we can then estimate the illumination change for each camera. To model the illumination changes we work on the same principle as the brightness transfer function outlined in Section 2.1. That is we assume that the percentage of pixels in background image $M_{i'}(x, y)$ with the brightness value less than $B_{i'}$ is equal to the percentage of image points seen in $M_i(x, y)$ of brightness less than or equal to B_i. And thus we use Equation (3) to compute $f_{i'i}$ and $f_{j'j}$ from Figure 2 as follows:

$$f_{i'i}(B_{i'}) = H_i^{-1}(H_{i'}(B_{i'})) \tag{7}$$

A similar approach was proposed by Grossberg et al[8], however their method does not consider significant background changes between images. As this mapping may not contain one-to-one brightness mappings we use linear interpolation to estimate unmapped regions. A sample illumination mapping can be seen in Figure 5. The mapping between j' and j is then calculated in the same way.

Once $f_{i'i}$, $f_{j'j}$ and the inter-camera CBTF cf_{ij} have been calculated using Equations (7) and (3) respectively objects can be mapped into the previous illumination conditions and objects in view i can be mapped to view j for comparison. Specifically, in order to compare two observations $O_{i',a}$ and $O_{j',b}$, their colours are converted to the corresponding colours in E_i and E_j, i.e. $\grave{O}_{i',a}(B_{i'})$ and $\grave{O}_{j',a}(B_{j'})$, using $f_{i'i}$ and $f_{j'j}$ respectively:

$$\forall B'_i, \grave{O}_{i',a}(B_{i'}) = f_{i'i}(O_{i',a}(B_{i'}))\forall B'_j, \grave{O}_{j',b}(B_{j'}) = f_{j'j}(O_{j',b}(B_{j'})) \tag{8}$$

Next $\grave{O}_{i',a}(B_{i'})$ must be converted to the illumination conditions of E_j, becoming $\hat{O}_{i',a}(B_i)$, using the learned inter-camera CBTF:

$$\forall B_i, \hat{O}_{i',a}(B_i) = cf_{ij}(\grave{O}_{i',a}(B_i)) \tag{9}$$

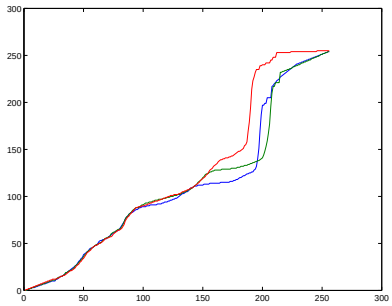


Fig. 5. Background illumination BTF from the blue channel of Site 3. Note the values on the x-axis (Day 2) corresponding to lower (darker) values on the y-axis (Day 1).

$\grave{O}_{i',a}(B_{i'})$ has now undergone transformation by a combination of $f_{i'i}$ and cf_{ij} as depicted in Figure 2.

Note that we have assumed so far that the CBTF contains only one-to-one colour relationships. However, in reality the mapping function obtained from the training set often contains cases of many-to-one colour correspondences due to incomplete ranges of colour values found in the training data. To address this problem, a nearest neighbour smoothing function is employed to smooth out the noisy peaks in the resulting histograms. Figure 6 shows an example of the process of converting a potential observation pair for comparison.

Once both $O_{i,a}$ and $O_{j,b}$ are converted to the same illumination conditions we can compare them directly using the Bhattacharya distance measure $D()$ and thus the similarity measure from Equation (1) can be defined as follows:

$$Similarity(O_{i,a}, O_{j,b}) = 1 - D(\hat{O}_{i,a}, \grave{O}_{j,b}) \tag{10}$$

Note that in order to compare two colour objects, we must apply this process to each of the three RGB channels. Thus the overall similarity measure becomes the mean of the similarity values obtained in all three channels.

2.3 Automatic Model Updating

The CBTF updating process described above is triggered automatically by the detection of an illumination change in each camera site. A background modelling approach such as [9] can be deployed to construct an empty background from a stack of frames containing foreground objects. From this automatically generated empty background region we can calculate the brightness histograms for the entry/exit region over a temporal sliding window. The brightness histograms are then compared against the brightness histograms of the background region from the previous period. Illumination change is detected when the difference between the brightness histograms is larger than a threshold.

3 Experiments

Two sets of experiments were carried out using challenging datasets collected from a distributed camera network. First, we compare the the proposed adaptive

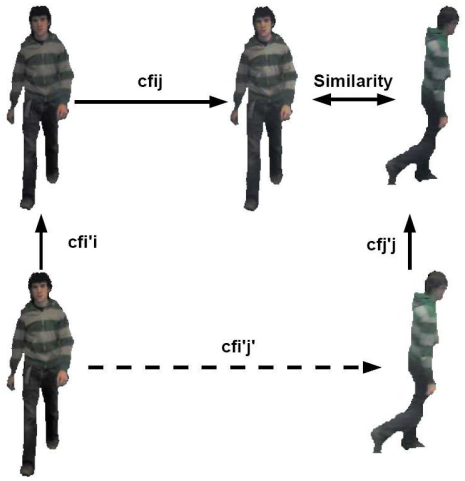
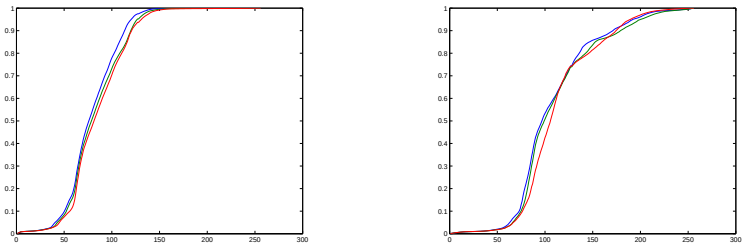


Fig. 6. Example of the conversion from the new illuminations (bottom row) to the old (top row). From here the image from camera i is converted to the illumination conditions of j for comparison using the similarity measure.

CBTF (A-CBTF) method against standard Bhattacharya distance and CBTF colour transfer [4] without temporal illumination change modelling. Secondly, we compare our method against current inter-camera colour compensation methods [1, 10]. In each of these experiments, the BTFs and CBTFs for each colour channel were estimated from a set of training pairs with known correspondences from Day 1 and tested using the observations in Day 2 which has different illumination conditions in each camera view (an example illumination change is shown in Figure 7). In each set of results we show both rank 1 and rank 5 results indicating the presence of the correct match as the highest and top 5 highest similarity scores respectively.



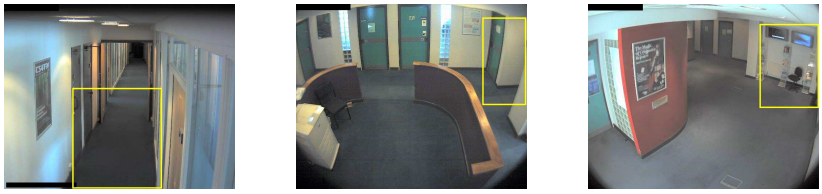
(a) Day 1 (b) Day 2

Fig. 7. Example RGB histograms of a single camera view on both days. Day 2 shows significant change in global illumination.

Datasets: We obtained two sets of data from inside an office building observed by three cameras. Example views are shown in Figure 8. The illumination conditions and colour quality vary between each views. Camera 1 displays a corridor scene where objects are periodically lit by spotlights causing darker regions in

the bottom part of a person's body. Camera 2 shows a shared space connecting several offices with fairly dim illumination. Camera 3 is placed in a foyer region where there is poor lighting in the back right region. Both datasets prove challenging as they contain sparse colour information and objects in similar clothing. The illumination conditions also vary greatly between the two data sets. The first data set, used for training the inter-camera CBTFs, was recorded on a cloudy afternoon. The second data set, used for testing, was recorded on a much brighter day. A single entry/exit region was determined in each camera to capture targets. The training and testing data were obtained from the entry/exit regions marked in yellow in Figure 8. In the training dataset, 15 individuals giving 45 entry/exit transitions were observed, and 20 individuals with 56 entry/exit transitions, were observed in the testing set.

(a) Site 1: Day 1 (b) Site 2: Day 1 (c) Site 3: Day 1



(d) Site 1: Day 2 (e) Site 2: Day 2 (f) Site 3: Day 2

Fig. 8. Sample frames from two days showing the differing lighting conditions between days in addition to the inter-camera illumination changes. The yellow boxes show the entry/exit zones.

Comparing Bhattacharya distance, CBTF, and Adaptive CBTF: Here we demonstrate that temporal illumination change modelling improves on the CBTF, which in turn is an improvement over Bhattacharya distance alone. Each observation was decomposed into its RGB and component histograms at each entry/exit region and compared against all other observations. For the Bhattacharya distance experiment, no colour mapping is performed. For the CBTF we use only the inter-camera CBTF learned from Day 1 (cf_{ij}) as an estimation of the colour changes between views in Day 2 ($cf_{i'j'}$). Figure 9 shows that the proposed method achieved an approximately 15% improvement in overall matching rate against Bhattacharya distance and CBTF. This validates our assumption that changes in illumination can be approximated using a linear combination of foreground and background changes. Example of object association results obtained using the three methods are shown in Figure 11. Figures 9(a) and 9(b) suggest that the improvement is significant for two camera pairs, whilst Fig-

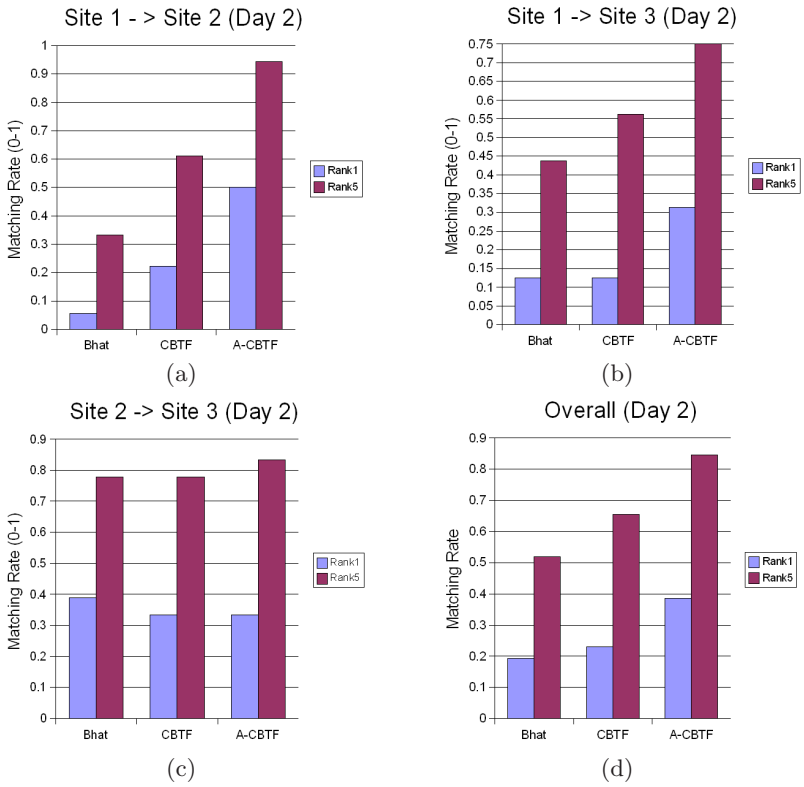


Fig. 9. Comparison of Bhattacharya distance, CBTF and A-CBTF with temporal illumination change modelling.

ure 9(c) show the three methods give a similar result for camera sites 2 and 3. This is because incidentally the illumination conditions at Sites 2 and 3 are similar, which causes the Bhattacharya distance result to be higher while the minor inaccuracies in the CBTF-based methods cause a slightly lower result.

Comparison with alternative approaches: In this experiment, our adaptive CBTF method is compared against other reported approaches. We have implemented the BTF subspace approach [1] and the Major Colour Spectrum Histogram (MCHR) approach [10], in which object colour histograms are equalised before being decomposed into major colours. Note, as there is no assumed knowledge of the relationship between cameras, our equalisation graph for the MCHR was based on a standard linear equalisation, whilst the graph in [10] was non-linear based on some rather arbitrary *a priori* knowledge. In addition, as the number of frames in which an object is captured passing through our entry/exit zones is low, the incremental MCHRs cannot be used. More critically though, as our model is designed for online processing, we have excluded their batch-based post matching integration part which cannot be performed online.

The results in Figure 10 show that the equalisation based MCHR does not cope well with this challenging data set. Although slightly better, the BTF sub-

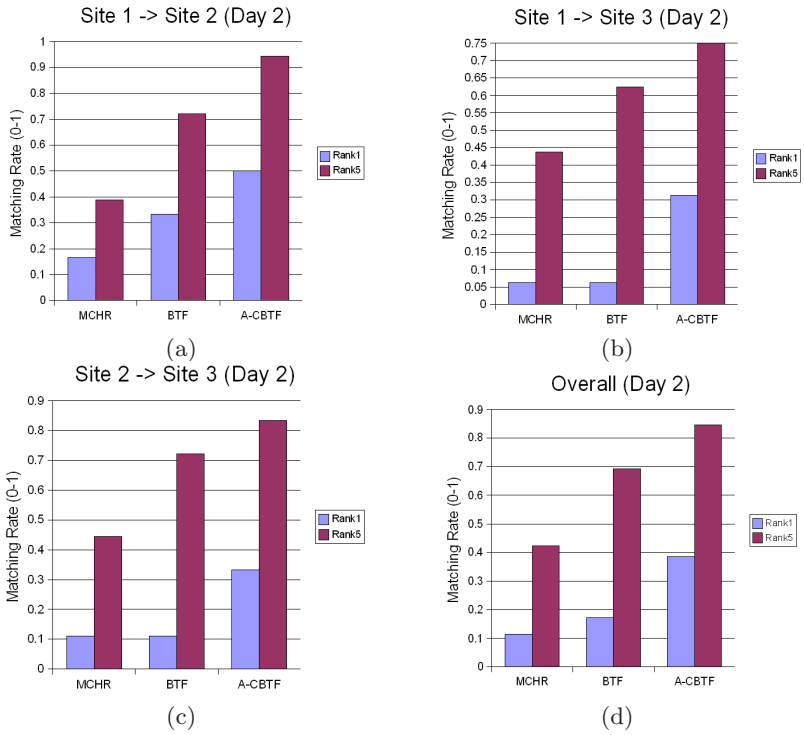


Fig. 10. Comparitive results from the MHCR-based method, the BTF subspace method and our adaptive CBTF method.

space approach suffers due to its inability to adapt to the difference between the illumination conditions changes over time. Overall our approach outperforms the two alternative methods by approximately 20% in rank 1 and 15% in rank 5.

4 Conclusions and Future Work

We have demonstrated that by modelling background illumination changes we can infer new brightness mapping functions between cameras from the original CBTF. In particular, by using background illumination we are able to estimate the changes on the foreground objects without the need for manual association of foreground objects each time these illumination conditions change, which would be required by other approaches. The datasets used provide a challenging test for object association due to the sparse colour information of the objects observed. Although our method produces relatively low matching rates its ability to adapt to new illumination conditions allows it to significantly outperform existing methods. In order to improve inter-object discrimination we plan to model the colour distribution of individuals. This would help us distinguish between objects with similar colour histograms but different colour layouts as can be seen in Figures 11(g) and 11(j). Currently our matching algorithm uses a brute force

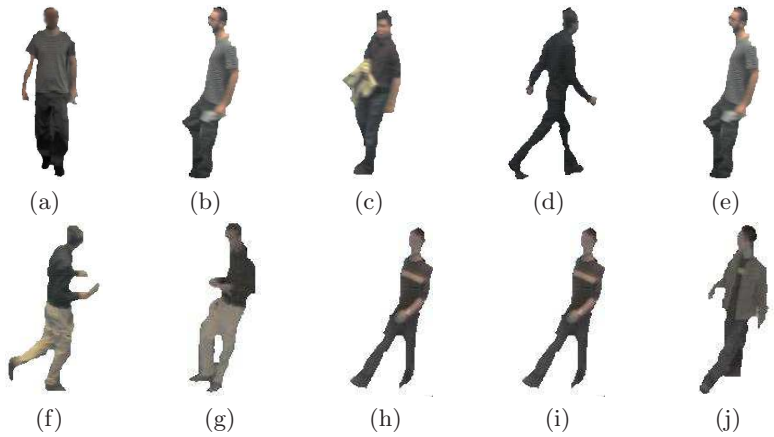


Fig. 11. (a) and (b): the same individual appeared at entry/exit regions 1 and 2 respectively.; (c): BTF(subspace) match; (d): MCHR match: (e): CBTF match (correct one). (f) and (g): A much more challenging case from due to the presence of similar coloured objects in the testing set. (h)-(j): all three methods found the wrong match.

approach to finding object correspondences. We also plan to incorporate temporal information [11] to reduce the number of initial correspondence hypotheses.

References

1. Javed, O., Shafique, K., Shah, M.: Appearance modeling for tracking in multiple non-overlapping cameras. In: CVPR. Volume 2. (2005) 26– 33
2. Ilie, A., Welch, G.: Ensuring color consistency across multiple cameras. In: ICCV. Volume 2. (2005) 1268 – 1275
3. Cheng, E.D., Piccardi, M.: Matching of objects moving across disjoint cameras. In: ICIP. (2006) 1769–1772
4. Prosser, B., Gong, S., Xiang, T.: Multi-camera matching using bi-directional cumulative brightness transfer functions. In: BMVC. (2008)
5. Gilbert, A., Bowden, R.: Tracking objects across cameras by incrementally learning inter-camera colour calibration and patterns of activity. In: ECCV. Number II (2006) 125–136
6. Makris, D., Ellis, T., Black, J.: Automatic learning of an activity-based semantic scene model. In: AVSS. (2003) 183 – 188
7. Li, J., Gong, S., Xiang, T.: Scene segmentation for behaviour correlation. In: ECCV. (2008)
8. Grossberg, M., Nayar, S.: Determining the camera response from images: What is knowable. In: PAMI. Volume 25. (2003) 1455 – 1467
9. Russell, D., Gong, S.: Minimum cuts of a time-varying background. In: BMVC. (2006) 809–818
10. Madden, C., Cheng, E.D., Piccardi, M.: Tracking people across disjoint camera views by an illumination-tolerant appearance representation. Machine Vision and Applications **18** (August 2007) 233–247
11. Makris, D., Ellis, T., Black, J.: Bridging the gaps between cameras. In: CVPR. Volume 2. (2004) II–205– II–210