Multi-camera Matching using Bi-Directional Cumulative Brightness Transfer Functions

Bryan Prosser, Shaogang Gong, and Tao Xiang
Department of Computer Science
Queen Mary, University of London
{bryan, sgg, txiang}@dcs.qmul.ac.uk

Abstract

The appearance of individuals captured by multiple non-overlapping cameras varies greatly due to pose and illumination changes between camera views. In this paper we address the problem of dealing with illumination changes in order to recover matching of individuals appearing at different camera sites. This task is challenging as accurately mapping colour changes between views requires an exhaustive set of corresponding chromatic brightness values to be collected, which is very difficult in real world scenarios. We propose a Cumulative Brightness Transfer Function (CBTF) for mapping colour between cameras located at different physical sites, which makes better use of the available colour information from a very sparse training set. In addition we develop a bi-directional mapping approach to obtain a more accurate similarity measure between a pair of candidate objects. We evaluate the proposed method using challenging datasets obtained from real world distributed CCTV camera networks. The results demonstrate that our bi-directional CBTF method significantly outperforms existing techniques.

1 Introduction

To re-establish a match of the same person over different camera views located at different physical sites is critical for object global behaviour analysis over disjoint multi-camera networks. To solve this object re-identification problem, one aims to match objects across camera views enabling the tracking of 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. While methods exist to address the problem of illumination change between camera views, none of them is able to deal with the problems inherent in real world data with low/varying image quality of very sparse colour information (see examples in Figure 1 collected from a public residential area CCTV network for our experiments). Our aim is therefore is to mitigate the effect on video data arising from these real world conditions.

Simple appearance based methods currently exist to handle the lighting condition changes between cameras. Javed et al [5] 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

Figure 1: Corresponding images of a person appearing in four entry/exit regions across three camera sites. Poor image quality and large variation in both colour and illumination pose serious problems for person re-identification even by an experienced human operator.

subspace to determine a correspondence. However, their method relies on training subjects with a good range of brightness values to give an accurate mean BTF. This implicitly assumes both extensive colour variations on object clothing and very large number of objects being sampled. Both assumptions are unlikely to be met given a limited time span in a real world scenario. Cheng et al [1] cluster colours into a subset of *major colours* and to alleviate the effect of illumination changes, they apply a histogram equalisation technique. As standard straight line (linear) equalisation is insufficient for modelling illumination changes in real world data, this approach assumes some arbitrary *a priori* knowledge of a suitable colour mapping function. This assumption is invalid in practice. Gilbert and Bowden [3] model inter-camera colour transformations using an incrementally updated transformation matrix. However, this method is computationally expensive as it requires thousands of objects to construct an accurate transformation matrix. A similar model was proposed in [4] without incremental learning. Instead, it requires a hardware calibration phase which is infeasible with camera installations of unknown camera parameters.

Most existing appearance based person re-identification methods rely on colour information exclusively. Recently more sophisticated appearance representations have been proposed to model texture and shape information in addition to colour. Wang et al [8] represent objects using histograms of oriented gradients that incorporate detailed spatial distribution of objects' colour across different body parts. Similarly Gheissari et al [2] segment human body into salient parts and combine colour and edgel histograms for appearance representation and person re-identification. However, both methods rely on objects having similar poses and observed in good quality data. It is evident from Figure 1 that under more realistic conditions texture and shape information are either non-existent or unreliable due to low image quality. The aforementioned methods thus stand little chance in solving the real world person re-identification problem.

In this paper we show that even given a sparse set of colour information a colour mapping function can be obtained and used to recognise individuals across camera views. Specifically, we propose to use a cumulative BTF (CBTF) as a more accurate representation of a set of BTFs compared to the subspace based method in [5]. Our approach involves an amalgamation of the training set before computing any BTFs in contrast with computing individual BTFs and then finding the mean [5]. This method maintains more of the colour information from the training set than the mean based approach. In addition we formulate a novel bi-directional matching criterion that allows us to assess the symmetry of a similarity measure used for comparing individuals in order to reduce false positives. This criterion is advantageous over both the uni-directional criterion used in most previous approaches and a conventional bi-directional one proposed in [6]. We evaluate the proposed method using challenging datasets obtained from real world CCTV cam-

era networks. The results demonstrate that our bi-directional CBTF method outperforms significantly existing approaches such as [5] and [6].

2 Brightness Transfer Function

Scene illumination varies between disjoint camera views, and in some cases within a single camera view. Thus, a vital stage in inter or intra-camera appearance based person re-identification is to mitigate the effect of such changes. Approaches have been proposed to find colour-to-colour correspondences between cameras and using these to create a colour mapping function known as Brightness Transfer Function (BTF). Javed et al [5] defined a non-parametric form of BTF that we will outline in this section.

Their method 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. Ideally this correspondence would be on the pixel level to ensure precise colour matches, but this is not possible due to differing object pose between views. Instead normalised histograms of RGB brightness values are used as these are more tolerant of changes in pose.

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 image I_i as follows:

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

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:

$$f_{ij}(B_i) = H_j^{-1}(H_i(B_i)) \tag{2}$$

with H^{-1} representing the inverted cumulative histogram. In order to produce a more accurate transfer function, multiple BTFs can be estimated. A BTF is typically calculated for each of a set of training pairs of observations and thus a set of BTFs $\{f_{ij}^1, f_{ij}^2, ...f_{ij}^N\}$ can be computed for cameras C_i and C_j given a training set of N observation pairs. An example of this can be seen in Figure 2 which shows a sample of BTFs taken from five individuals given 5 pairs of appearances in two different cameras. From this set a mean BTF $\overline{f_{ij}}$ can be produced to incorporate all of the training set information. This mean BTF can then be used to match objects by transforming testing observations from one camera to another, or by comparing testing BTFs against this mean BTF in a subspace as proposed in [5].

3 Cumulative Brightness Transfer Functions

Mean BTF based methods rely on having a consistent set of coloured individuals to accurately model the BTF. Taking the mean of a set of BTFs actually removes vital colour

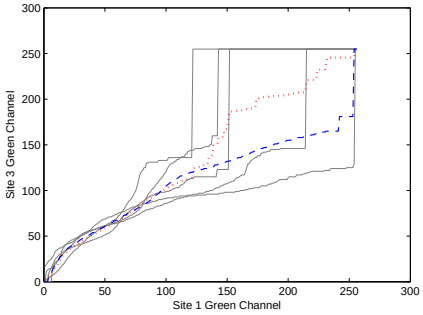


Figure 2: Five example BTFs (coloured grey) in the green channel taken between Site 1 and Site 3 from Scenario 1 (see Figure 4) used in our experiments. The sharp increase in the gradient of the lines is due to a lack of high end colour values in the data. The mean BTF is displayed in red (dotted) and the proposed CBTF in blue (dashed).

information that may only be contained in a small subset of the training data. For example, if most of the training data consists of dark clothed individuals and one single person wearing a bright blue shirt, the averaging process will remove most of the useful bright colour information from this individual, which is under sampled in the training set. To combat this we propose a cumulative approach to averaging sets of training BTF. Instead of computing a BTF for each training pair we propose an accumulation of the brightness values of the whole training set before the BTF computation. The cumulative histogram $\hat{H}_i$ of N training samples in camera view i can be computed from the brightness values $B_1,...,B_m,...,B_{256}$ as:

$$\hat{H}_i(B_m) = \sum_{k=1}^{m} \sum_{L=1}^{N} I_L(B_k). \tag{3}$$

Note that this cumulative histogram must then be normalised by the total number of pixels in the training set to allieviate the effect of size difference between views. After obtaining this single cumulative histogram using all image pairs in a training set, we can then use the histogram to compute a cumulative BTF (CBTF) as follows

$$cf_{ij}(B_i) = \hat{H}_j^{-1}(\hat{H}_i(B_i)) \tag{4}$$

The key advantage of CBTF over a standard mean BTF is that brightness values that are not common in the training set are still preserved. As a result uncommon brightness values in the training data can be more accurately mapped between cameras. This advantage is demonstrated in Figure 2. It can be seen that the mean BTF is affected by the lack of bright colour values in some of the BTFs which causes a premature rise in both the original BTFs and the mean BTF. In contrast, our CBTF retains the colour information of all the initial training BTFs and produces a more accurate colour mapping function.

4 Re-Identification using Bi-Directional CBTF

A camera network has m cameras $C_1,...,C_m$ all of which are assumed to have no-overlapping views. Unlike the approach in [5] which considers whole camera 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 $E_1, ..., E_g$. These entry/exit zones can be either manually defined or automatically learned [7]. 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 $Q^{j,b}_{i,a}$ 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 $Q^{j,b}_{i,a} \in s$, 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:

$$s = \arg\max_{s} \prod_{Q^{j,b}_{i,a} \in s} Similarity(O_{i,a}, O_{j,b}) \tag{5}$$

where $Similarity()$ is the similarity between $O_{j,b}$ and $\hat{O}_{i,a}$. To compare two observations $O_{i,a}$ and $O_{j,b}$, the colours of $O_{i,a}$ are converted to the corresponding colours in E_j using $cf_{ij}(\,)$ such that

$$\forall B_i, \hat{O}_{i,a}(B_i) = cf_{ij}(O_{i,a}(B_i)) \tag{6}$$

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 histogram. Until now we have considered only a single colour channel. In order to compare two colour objects, we calculate and apply the CBTF to each of the three RGB channels. Thus the overall similarity measure is the mean of the similarity values obtained in all three channels.

The similarity between $O_{j,b}$ and $\hat{O}_{i,a}$ is calculated using $1-$ Bhattacharya distance $Sim(\hat{O}_{i,a}, O_{j,b})$. This process can be repeated for the transfer in the opposing direction by transferring $O_{j,b}$ into the colours found in E_i and thus comparing $O_{i,a}$ and $\hat{O}_{j,b}$ using $Sim(O_{i,a}, \hat{O}_{j,b})$. The transfer functions, and the resulting similarity score, are subject to some differences depending on direction. This means that one direction may result in better matching scores or a combination of the two may increase the result. However, this cannot be determined before the results are obtained. In order to utilise the additional information from this bi-directionality the following methods are considered:

- *Mean:* Assuming that the similarity values for each direction give close numerical results an average of the two values are used to estimate the overall match.
- *Maximum:* Taking the highest value of the two as the matching result ensures that if one direction produces a better matching score it will be selected, but may increase the chances of false positives.

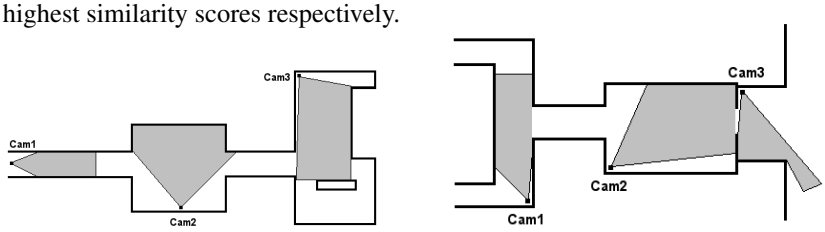
- *Minimum:* Taking the smaller of the two values assumes that both values will be high enough to qualify as a match but selects the lower each time to try and reduce false positives and thus the overall matching rate.

- *Symmetry Ratio Weighting (SRW):* Assuming that a correct match will produce a higher and more symmetric $Sim()$ score for each direction and an asymmetric score would indicate an incorrect match. We propose a adaption of the similarity score presented in [6] to incorporate the mean of the $Sim()$ values as follows:

$$Similarity(O_{i,a}, O_{j,b}) = \left(\frac{Sim(\hat{O}_{i,a}, O_{j,b}) + Sim(O_{i,a}, \hat{O}_{j,b})}{2} \right) \left(1 - \frac{Sim_{max} - Sim_{min}}{Sim_{max} + Sim_{min}} \right) \tag{7}$$

5 Experiments

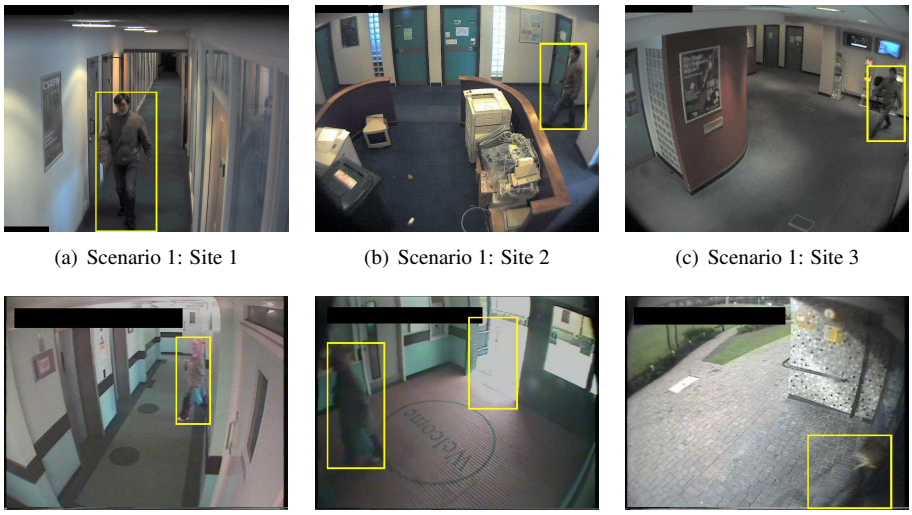
Three sets of experiments were carried out using challenging datasets collected from two distributed camera networks of real world scenarios. First, we compare the proposed CBTF and the mean BTF using a uni-directional transformation in order to demonstrate that the estimated mapping function using CBTF is more accurate. Second, we compare this uni-directional CBTF approach with the proposed bi-directional CBTF approach to evaluate the effect of the proposed bi-directional similarity measure. Finally, we compare our bi-directional CBTF method against alternative approaches from [5, 6]. In each of these experiments, the BTFs and CBTFs for each colour channel were estimated from a set of training pairs with known correspondences. In each set of results we show both rank1 and rank5 results indicating the presence of the correct match as the highest and top 5 highest similarity scores respectively.



(a) (b)

Figure 3: (a) Scenario 1 camera configuration. All cameras are mounted indoors. (b) Scenario 2 camera configuration. Cameras 1 & 2 are indoors whilst camera 3 is outdoors.

Datasets: The first scenario (referred as Scenario 1) is inside an office building observed by three cameras. The topology of this camera network is shown in Figure 3(a) with example views shown in Figure 4(a)-(c). The illumination conditions and colour quality vary between these 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 making it a good spot to test potential algorithms. 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. Our dataset consists of synchronised videos recorded simultaneously from 3 different cameras. In this dataset, 15 individuals giving 45 entry/exit transitions were used in the training phase, and the remaining 20 individuals with 51 entry/exit transitions, were used in testing.



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

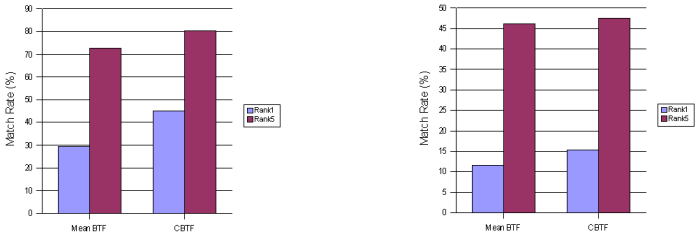
(d) Scenario 2: Site 1 (e) Scenario 2: Site 2 (f) Scenario 2: Site 3

Figure 4: Sample frames from two scenarios: the same person reappeared in different camera sites in each scenario. The yellow boxes show the entry/exit zones. The different camera views in both scenarios undergo significant changes in both illumination and pose.

The second experimental scenario (referred as Scenario 2) was obtained from both inside and outside a residential building. The camera topology is depicted in Figure 3(b). Camera 1 shows a foyer scene with relatively rich colours and good illumination. Camera 2 shows a large variation in illumination from right to left due to the presence of an outside door on the right hand side of the view. We thus capture data from the entry/exit region on each side of this camera view. Camera 3 captures objects entering the building. Due to the stark differences in illumination and colour between the 4 entry/exit regions, this is an even more challenging dataset than that from Scenario 1. From this dataset 63 and 78 entry/exit transitions were used in training and testing respectively.

Mean BTF vs. CBTF: In order to show that the CBTF provides a better estimation of the colour mapping between entry/exit regions we use a uni-directional comparison using the Bhattacharya distance as similarity measure. For each individual we converted their RGB histograms to the target entry/exit region colour space. They were then compared against all individuals observed in this region. Figure 5(a) shows an approximate 15% improvement in matching rate when compares CBTF with the mean BTF. In Figure 5(b), it can be seen that although both methods are affected by the harsher illumination and colour differences in Scenario 2, the CBTF is still a better approximation of the mapping function. An example of the colour mapping using mean BTF and CBTF can be seen in Figure 6.

Bi-directional vs. uni-directional: In this experiment, we demonstrate the differences in results between the two possible directions of colour transfer and that by adding a comparison method to the two directions we can mitigate the effect of the differences in their value. Figure 7 shows that only using the single direction matching can produce different results depending the on the direction chosen, of which the dominant direction may differ between data sets as show or even between individual objects. Of the bi-directional measures tested, the minimum value clearly indicates that by attempting to



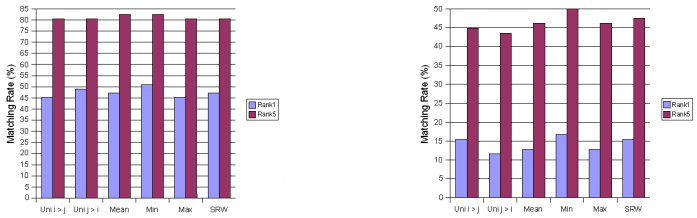
(a) Scenario 1 (b) Scenario 2

Figure 5: Compare CBTF with mean BTF using uni-directional similarity matching.

(a) (b) (c) (d)

Figure 6: (a) Original frame from Scenario 1 entry/exit point 3. (b)The same individual in entry/exit point 2. (c) Original frame in (a) mapped using CBTF which results in a correct matching. (d) Original frame mapped mean BTF which results in a wrong matching. Note the mean BTF inaccurately maps the higher brightness values found in the white top.

remove false positive matches from each direction we can extract the information from the more accurate direction and also improve the overall match rate. The improvement made by the SWR approach was lower than expected due to the sparse colour distribution in the datasets resulting in less variation in the symmetry values.

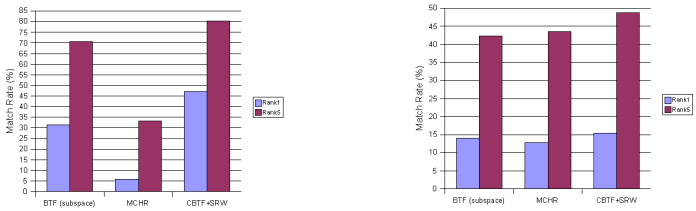


(a) Scenario 1 (b) Scenario 2

Figure 7: Comparing bi-directional and uni-directional matching using CBTFs.

Comparison with alternative approaches: In this experiment, our bi-directional similarity ratio weighted CBTF method is compared against other reported approaches. We have implemented the BTF subspace approach [5]. The other approach used for comparison is based on the Major Colour Spectrum Histogram (MCHR) approach [6], 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 [6] 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.



(a) Scenario 1 (b) Scenario 2

Figure 8: A comparison of the matching success rates of the BTF subspace method, MCHR colour conversion and the proposed Bi-Directional CBTF method.

The results from Scenario 1 (Figure 8(a)) show that the MCHR is harshly affected by both illumination changes and visual appearance changes of objects. The BTF subspace approach performs better than MCHR in both rank 1 and rank 5 scores. In comparison, the performance of our method is significantly better than both. In particular, the bi-directional CBTF method obtains more than 80% match rate in the rank 5 comparison and an almost 20% increase in rank 1 matching rate over the BTF subspace method, demonstrating its clear superiority in overcoming both illumination changes and changes in the visual appearance of objects. Due to the challenging circumstances in the Scenario 2 dataset (Figure 8(b)), all 3 methods produce low rank 1 results but our method shows some improvement in the rank 5 accuracy. Figure 9 shows an example of matched and unmatched objects using the three different approaches. The transfer from the faded red in Figure 9(a) to the higher brightness values in Figure 9(b) is better defined in our CBTF method thus giving a correct match. Figures 9 (f)-(j) show an extremely challenging case for appearance based re-identification where all three method failed.

(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

Figure 9: (a) and (b): the same individual appeared in Scenario 1 entry/exit points 3 and 2 respectively.; (c): BTF(subspace) match; (d): MCHR match: (e): CBTF match (correct one). (f) and (g): A much more challenging case from Scenario 2 due to self occlusion of the bag and poor segmentation. (h)-(j): all three methods found the wrong match.

6 Conclusions and Future Work

We have shown that an accumulative representation prior to calculating brightness transfer functions improves model estimation when a full range of brightness values is not observed or unavailable in the training data. We have also demonstrated the advantage of a bi-directional CBTF re-identification approach in ensuring the colour mapping information from both directions is considered therefore reducing false positives. The datasets presented in this work pose challenging circumstances for object re-identification. Although the matching rates show improvement over alternative approaches, there are unsolved problems. Currently our method uses a brute force approach to re-identification by comparing a target unknown individual with all known individuals in all cameras. We plan to add temporal links to the individuals, such as camera transition time [7], which has been shown to improve tracking results [5, 3]. We also plan to develop an automated online updating method. This method would allow us to include new colour matches for previously unseen brightness values within the CBTF. Additionally, in order to further distinguish between individuals of similar appearance we plan to incorporate spatial information about a person's appearance. The spatial information could also be used to aid the matching of occluded objects as seen in Figures 9(f) and 9(g).

Acknowledgement: The authors shall thank Ray Stead and his colleagues of the Portsmouth City Council and Nick Hewitson of SmartCCTV for their assistance in collecting CCTV data.

References

[1] E. D. Cheng and M. Piccardi. Matching of objects moving across disjoint cameras. In *ICIP*, pages 1769–1772, October 2006.

[2] N. Gheissari, T. Sebastian, P. Tu, and J. Rittscher. Person reidentification using spatiotemporal appearance. In *CVPR*, volume 2, pages 1528 – 1535, 2006.

[3] A. Gilbert and R. Bowden. Tracking objects across cameras by incrementally learning inter-camera colour calibration and patterns of activity. In *ECCV*, number II, pages 125–136, 2006.

[4] A. Ilie and G. Welch. Ensuring color consistency across multiple cameras. In *ICCV*, volume 2, pages 1268 – 1275, 2005.

[5] O. Javed, K. Shafique, and M. Shah. Appearance modeling for tracking in multiple non-overlapping cameras. In *CVPR*, volume 2, pages 26– 33, 2005.

[6] C. Madden, E. D. Cheng, and M. Piccardi. Tracking people across disjoint camera views by an illumination-tolerant appearance representation. *Machine Vision and Applications*, 18:233–247, August 2007.

[7] D. Makris, T. Ellis, and J. Black. Bridging the gaps between cameras. In *CVPR*, volume 2, pages II–205– II–210, 2004.

[8] X. Wang, G. Doretto, T. Sebastian, J. Rittscher, and P. Tu. Shape and appearance context modeling. In *ICCV*, pages 1–8, October 2007.