A Database for Person Re-Identification in Multi-Camera Surveillance Networks

Alina Bialkowski, Simon Denman, Sridha Sridharan, Clinton Fookes
Image and Video Research Laboratory
Queensland University of Technology, Brisbane, Australia
Email: {alina.bialkowski, s.denman, s.sridharan, c.fookes}@qut.edu.au

Patrick Lucey
Disney Research Pittsburgh
Pittsburgh, PA, USA, 15213
Email: patrick.lucey@disneyresearch.com

***Abstract*—Person re-identification involves recognising individuals in different locations across a network of cameras and is a challenging task due to a large number of varying factors such as pose (both subject and camera) and ambient lighting conditions. Existing databases do not adequately capture these variations, making evaluations of proposed techniques difficult. In this paper, we present a new challenging multi-camera surveillance database designed for the task of person re-identification. This database consists of 150 unscripted sequences of subjects travelling in a building environment though up to eight camera views, appearing from various angles and in varying illumination conditions. A flexible XML-based evaluation protocol is provided to allow a highly configurable evaluation setup, enabling a variety of scenarios relating to pose and lighting conditions to be evaluated. A baseline person re-identification system consisting of colour, height and texture models is demonstrated on this database.**

I. Introduction

In a surveillance network, it is often desirable to be able to recognise and track people as they move through the environment. In a single camera view, this can be achieved through object tracking techniques, however, in a large space with multiple non-overlapping cameras where it is not certain which path people will take, appearance matching methods must be applied to re-identify an individual as they move between cameras. This problem is termed person re-identification, and involves recognising an individual in different locations across a network of cameras, typically assuming that individuals wear the same clothing between sightings, as represented in Figure 1.

Despite the assumption that people within the environment have the same appearance from camera to camera, several complexities which arise from the environment make this a challenging problem. These factors include:

1) subjects will often only be visible at low resolution;
2) subjects may appear at different poses and viewpoints (e.g. front-on or side-on) as they move through the camera network;
3) the environment often contains many different lighting conditions, altering the appearance of people in the space;
4) and subjects may be partially occluded (e.g. by bags or other people).

In such conditions, traditional biometrics such as face, iris or gait generally cannot be used. Instead, models which characterise the overall appearance of a person, or models which consist of a collection of local descriptors are used. Such models are often termed “soft biometrics" [1] and are defined as characteristics which can be used to describe, but not uniquely identify an individual. Soft biometrics include traits such as height, body build, gender, ethnicity, and characteristics which may change more frequently such as clothing colour. Using such features, we can detect if a given person has been previously observed elsewhere in a network of cameras, or search for an individual in a camera network.

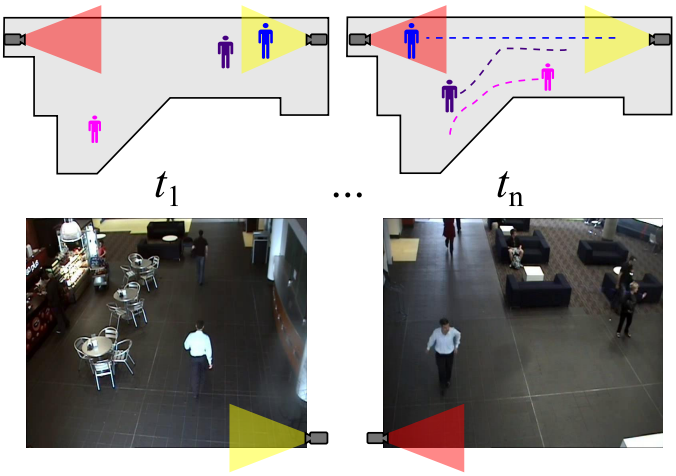


Fig. 1: A scene at two time instants, t_1 and t_n, is represented, with the coloured people representing different identities. Person re-identification seeks to recognise the identity of a person as they move between different locations, given a set of previously observed people. For example, the blue person visible from the yellow camera at t_1, later appears in the red camera at t_n. A person re-identification system should be able to reconcile this identity despite the change in appearance in the acquired video frames.

To evaluate models for person recognition and re-identification, a dataset is required which consists of multiple cameras, in which the subjects appear in different poses, viewing angles and lighting conditions. Due to the limitations of existing databases that either contain only still images (i.e. VIPER [2]), few camera views (i.e. ETHZ [3], PETS 2006 [4]), highly controlled conditions (i.e. CASIA [5]), or a

lack of sufficient frames per subject (i.e. i-LIDS MCTS [6]), a new database is proposed. This new database consists of 150 people, with an average of over 400 frames per person spanning up to eight camera views in challenging surveillance conditions. A flexible XML-based evaluation protocol is provided to allow for a highly configurable evaluation setup, enabling a variety of scenarios relating to pose and lighting conditions to be evaluated.

This new dataset provides a platform from which to answer questions such as:

- What features are best for recognising the identity of a person in low resolution footage across different camera views, illumination conditions and with variable pose?
- How much data is necessary to build a sufficient model of a person?
- How does data from multiple views impact performance?
- Can details about pose be used to improve performance?

We demonstrate the utility and flexibility of the proposed database by using it to answer these questions with a baseline person re-detection system consisting of colour, height and texture features.

The remainder of this paper is structured as follows: Section II covers related work in the field of person re-identification and the existing databases used in evaluations; Section III describes our new multi-camera surveillance database; Section IV describes the baseline models which are used to demonstrate the utility of our database, followed by results in Section V, and conclusions in Section VI.

II. Related Work

A. Person re-identification

In a surveillance environment, traits that can be observed at a greater range are desirable, and such traits should be invariant to view and to lighting conditions.

Colour features are commonly used to model appearance and can be used to encode information about a person's clothing, hair and skin colour. They are popular for use in surveillance as they are mostly view invariant and can be sensed at a far distance from a camera. The most common method of utilising colour information is through histograms. Position information can be incorporated by splitting the person into parts (e.g. in [7], [8], histograms are extracted for the head, torso and legs) which allows matching based on colour and distribution. A more advanced approach such as the Mean Riemannian Covariance Grid (MRCG) [9] can better provide colour and spatial information.

Histograms allow for some degree of variation in colour caused by illumination, as a range of colours are allocated to each histogram bin. A "soft" binning approach [10] can be applied to further compensate for illumination changes and prevent the case where similar colours are allocated to different bins. In soft histogram binning, a pixel colour value is allocated to multiple bins, weighted according to the pixel value's proximity to the centre value of each bin.

Illumination changes between cameras can be compensated for using image based transformations [11], or a brightness transfer function between cameras [12] can be learned with prior training. Culture colours [13], which are a set of 11 colours recognised by most cultures (black, blue, brown, green, grey, orange, pink, purple, red, yellow, white), can also be used as they are less prone to variation across cameras.

Some approaches to person re-identification use texture based features or interest points to match people between cameras. Hamdoun et al. [14] use interest points to detect people across different views, however the method is only evaluated on a dataset of 10 subjects across 2 camera views. Gheissari et al. [15] use a decomposable triangulated graph model to segment a person into six horizontal strips and for each strip, extract HSV colour information, and edgels which encode edge orientation (vertical or horizontal), and the colour change across the edge. This method is evaluated on a 44 subject dataset across 3 cameras views (consisting of mostly frontal frames of a person).

Other methods for person re-identification combine colour and texture features, and aim to extract texture features which are view independent. Bazzani et al. [16] proposed a person descriptor which includes a global HSV histogram, an 'average' texture of the person and a a set of recurring textural motifs within the subject. This work was extended by Farenzena et al. [17] by using a symmetry-driven approach to extract features, and by including Maximally Stable Colour Regions (MSCRs) [18] in the appearance models. Bak et al. [19] proposed appearance models based on Haar-like features and dominant colour descriptors. The most invariant and discriminative signature was extracted using the AdaBoost algorithm. Schwartz et al. [20] proposed a large feature set consisting of texture, edge and colour information projected into a low-dimensional discriminant latent space using Partial Least Squares (PLS) reduction. The PLS scheme is shown to outperform PCA and SVM approaches.

While these methods have demonstrated applicability in the datasets provided, it is uncertain how they would perform in different conditions, as the datasets do not allow for different evaluation conditions. Even though many of the discussed features are designed to be view and illumination tolerant, not all the datasets are able to show that this is the case, and none are able to show *how* the models are affected by viewing angle or illumination. Also, many approaches only look at the single image case, which is unrealistic in a surveillance network, as video is captured and available for use to perform foreground segmentation and allows for better selection of frames to use in the model.

B. Existing datasets

To date, researchers have used a variety of data sources to evaluate their models. Existing tracking databases have been used (e.g. [8] used a subset of PETS2006 [4]); the VIPeR (Viewpoint Invariant Pedestrian Recognition) database [2] has been used extensively (see [2], [17], [21]–[23]); some have used the ETHZ [3] and i-LIDS [6] databases; while others have simply captured their own data (e.g. [15], [19]).

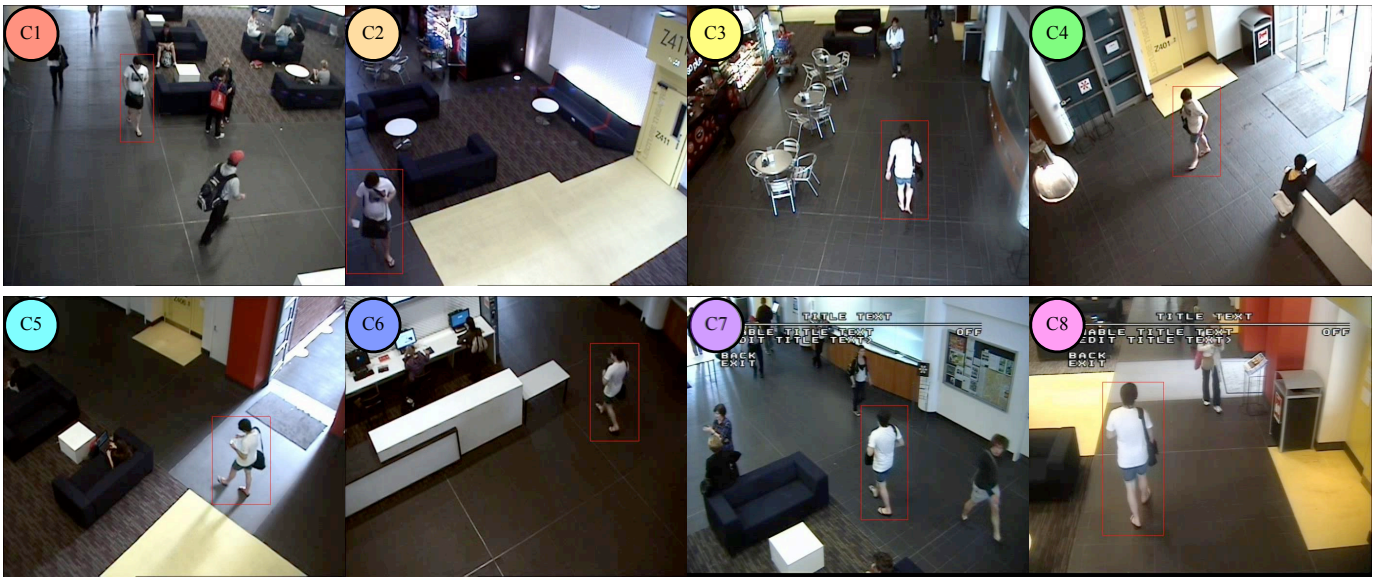


Fig. 2: Example video frames from each of the eight cameras (C1 to C8) of our database. A subject dressed in a white shirt, marked with a red bounding box, is shown in each of the cameras, highlighting the significant appearance variations (pose, viewpoint, illumination) as the subject moves through the camera network.

While these databases have their merits, it is difficult to compare and evaluate person re-identification models in real environments due to the lack of a suitable database. While PETS 2006 (and similar tracking databases), VIPeR, ETHZ, and i-LIDS are public data sets, they are limited for soft biometric applications. Tracking data sets typically consist of few cameras and a small number of distinct subjects for whom there is a suitable amount of footage for a soft biometric evaluation (PETS 2006 has four cameras of which only three are suitable), VIPeR is limited to a single image of each pedestrian from two viewpoints, ETHZ is captured from a moving stereo rig, and hence only captures similar (mostly frontal) viewing angles of a person, and the annotated component of i-LIDS only contains up to four images per person. While databases used in gait recognition research often contain a larger number of subjects and camera angles (e.g. the CASIA database [5] contains over 100 subjects observed from 11 cameras), they are captured in highly controlled conditions, very dissimilar to a typical surveillance environment.

III. The Multi-Camera Surveillance Database

The multi-camera surveillance database [1] was captured from an existing surveillance network, to enable the evaluation of person recognition and re-identification models in a real-life multi-camera surveillance environment. The database consists of 150 people moving through a building environment, recorded by eight surveillance cameras. Each camera captures data at 25 frames per second, at a resolution of 704×576 pixels, and is calibrated using Tsai's method [24]. An example image from each camera is shown in Figure 2, with the approximate camera placement and orientation displayed in Figure 3. The placement of cameras is a real-life surveillance setup, and cameras have been placed to provide maximal coverage of the space (with some overlap) and observation of the entrances to the building.

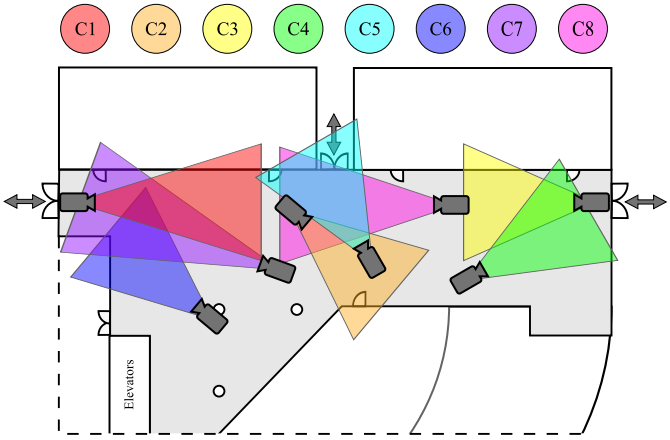


Fig. 3: Approximate camera placement and orientation in the Multi-Camera Surveillance Database. The three entrances to the building are indicated with arrows.

The database was collected in an uncontrolled manner, so subjects can travel any route through the building. Thus, the vast majority of subjects will only pass through a subset of the camera network and that subset varies from person to person. This provides a highly unconstrained environment in which to test person re-identification models. From Figure 2 and 4, it can be seen that there is varied lighting across

[1] Available from http://eprints.qut.edu.au/53437/ or by contacting the authors

Fig. 4: Example annotations of four subjects from the Multi-Camera Surveillance Database at different locations in the camera network, where S represents the subject ID and C represents the camera number.

the different camera views, and that subjects are observed from different angles as they move through the network. To enable a consistent evaluation in such conditions, a coarse bounding box indicating the location of the subjects has been annotated (every 20th frame was annotated and intermediate frame locations were interpolated). The frames are recorded from when the subject enters the building through one of the three main doorways visible in Camera 4, Camera 7 and Camera 5/8, until they leave observation either through exiting the building or entering a lecture theatre. Any frames which are significantly occluded, have been omitted. Examples of the annotated subjects are shown in Figure 4

XML files are used to store information about the database to enable different evaluations to be easily performed based on which subset of the database fits the desired criteria. For each subject, an XML file is used to summarise the camera views and frame information which can be used to select subjects which fit the desired evaluation conditions (e.g. only subjects that exist in specific cameras or locations can be selected). The overall database is also summarised in an XML file, which provides information on the camera calibration data for each subject. Zones of interest can be specified to further filter the person annotations, allowing for additional conditions to be evaluated (i.e. lighting changes can either be reduced or emphasised by only considering certain scene areas).

The database provides great flexibility in the possible evaluations that can be carried out due to the variations captured by the eight cameras. It can be used for traditional biometric identification and verification tasks, as well as the tracking person re-detection simulated by Synthetic Recognition Rates [2].

IV. Person Models

In this work, we consider colour, height and texture models for a person. The overall evaluation procedure and the steps to acquire our baseline models is displayed in Figures 5 and 6.

For all models outlined within this section, a motion segmentation algorithm [25] is used to separate the subject from the background. After extracting the foreground regions (i.e. pixels belonging to the person), the person is divided into head, torso and legs parts through horizontal projection and image gradient analysis as described in [26]. Example output from this process is shown in Figure 7.

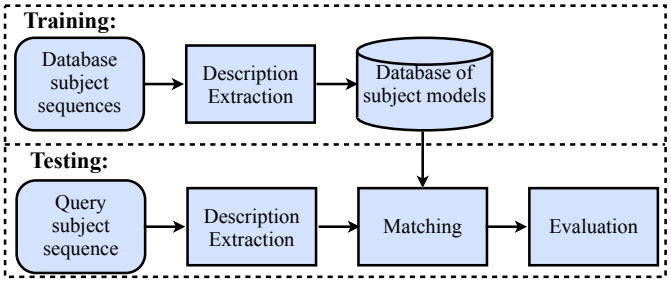


Fig. 5: Person re-identification system evaluation flowchart

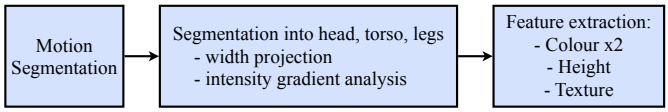


Fig. 6: The steps involved in extracting a description of a person in our baseline system

A. Colour Models

Colour information of a person is extracted by computing histograms of their head, torso and leg regions. For each of the three regions, a soft histogram of the full colour space is calculated as well as a histogram of the culture colours [13], resulting in two colour soft biometric models (soft histogram and culture colour histogram). A moving average of each histogram is calculated to incorporate multiple frames into the model.

In the soft histogram, variation in colour across different cameras is reduced through the soft-binning, where each pixel colour value is assigned to multiple bins based on its proximity to the centre of each bin. This means that samples which lie on a bin boundary, where there is greater uncertainty, are split

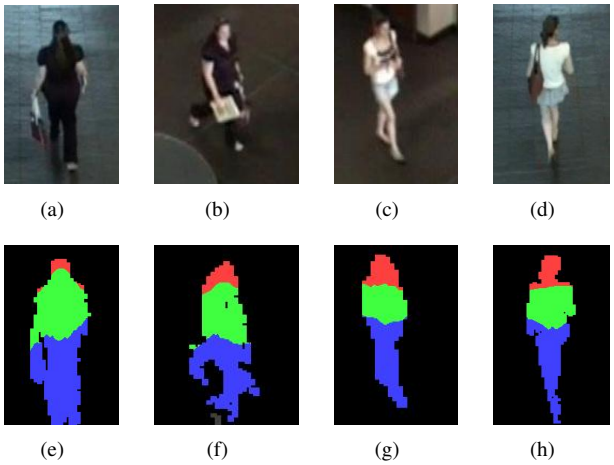


Fig. 7: Segmenting a person into head, torso and leg regions (coloured in red, green and blue respectively). The top row shows the input colour images, the bottom row shows the segmented silhouettes.

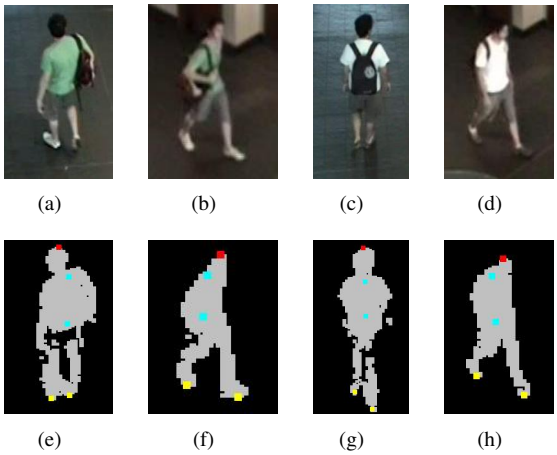


Fig. 8: Detecting the head, neck, waist, and feet. The top row shows the colour input image and the bottom row shows the corresponding silhouette with the detected points overlaid. The head points are shown in red, feet shown in yellow, and median position of the waist and neck divisions shown in cyan.

more evenly and prevents very similar colours from being wholly allocated to different bins.

The culture colour model quantises the image into 11 colours (black, brown, grey, red, orange, yellow, green, blue, purple, pink, white), with the aim of transforming the colours into a space less affected by illumination variations. To convert the image into its corresponding culture colour image, Gaussian mixture models (GMMs) were trained to represent each of the 11 culture colours from a set of small image patches (each containing a single culture colour). Each foreground pixel of a person is then classified into the culture colour with the greatest likelihood, and then the histograms are computed.

The histograms are normalised to sum to 1, ensuring invariance to the number of images used to build the model and the size of those images, and are compared using the Bhattacharya coefficient. When comparing colour models for two people, the similarity score is taken as the average of the three histogram region (head, torso, legs) comparisons.

B. Height Model

The height of a person is used as a simple descriptor as it is most view invariant. Other dimensions (width and depth) are dependent on the camera angle and a person's pose.

Heights are calculated using the detected positions of the head, torso and legs (which are converted into a real world coordinate scheme using camera calibration), and we use a soft histogram approach as described in [27]. Figure 8 shows an example of the located head and feet points, and the points used to divide the subject into head, torso and legs.

C. Texture Model

To model the texture information of a person, we calculate local binary patterns (LBPs) [28]. The LBP is an excellent texture descriptor for its invariance to illumination, and can also be made to be rotation invariant. In this work, we use an LBP model consisting of 8 points with a radius of 1 pixel, and a single texture model is extracted for the whole person, resulting in a feature vector of size 256. The LBP calculation procedure, from which the histograms are built, is shown in Figure 9.

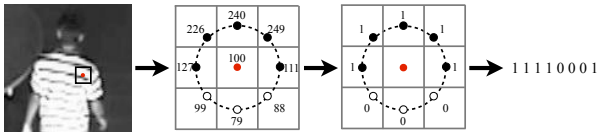


Fig. 9: Calculating the LBP feature vector

D. Fusion

As each model (colour, height, texture) forms a weak classifier, they can be fused together to take advantage of the complementary information of each model. We apply a weighted summation of the models, so that the overall match between two people i and j is:

$$M(i,j) = \sum_{n=1}^{4} w_n \times M_n(i,j), \tag{1}$$

where w is the weight applied to model n (soft histogram colour model, culture colour model, height model and texture model); and $M_n(i,j)$ is the matching score for model n, between person i and j.

V. Results

To demonstrate the utility of the proposed database, we investigate how the baseline soft biometrics are affected by a variety of factors captured by this database. We present the results for the following evaluations:

1) effect of the number of frames considered in the models
2) effect of viewing angle
3) effect of the number of camera views considered in the models

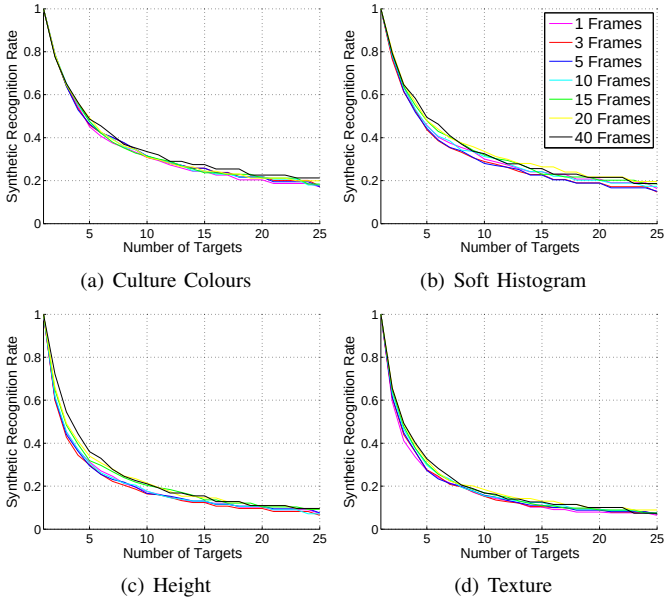
Results are presented using Cumulative Matching Characteristic (CMC) curves, which represent the probability of finding the correct match in the top x matches, and Synthetic Recognition Rate (SRR) curves which represent the probability that any of the y best matches is correct, as proposed in [2]. Note that the number of subjects present in each evaluation is not consistent as only subjects that match the criteria set out for the given evaluation are used. As the database is unconstrained, different numbers of people appear in different cameras, leading to this variation.

A. *Effect of number of frames considered in the model*

As a person moves through the environment, their sensed appearance will change according to the camera and ambient conditions. By considering more frames we expect more of this variation to be incorporated in the models. Results for this evaluation are presented using SRR instead of CMC curves, as they better represent the difference with the variable number of subjects (as we increase the number of frames for modelling, less subjects are available which fit this criteria in the database). In Figure 10 and Table I, a slight improvement is observed when considering more frames in the models (SRR values generally increase as more frames are considered, with best performance always obtained using 20 or 40 frames). Sometimes a slight decrease is observed which may be caused by noise being incorporated in the models, for example due to segmentation errors or strong lighting variations. While generally only a small improvement is gained, having a dataset with many frames allows for motion segmentation to be performed, so only pixels belonging to a person will be incorporated in the models. Having multiple frames available for modelling a person is more representative of a realistic scenario (surveillance is captured as video), and with more frames available, criteria can be applied to filter out frames detected to be of poor quality (e.g. poor segmentation/illumination as in Figure 13 (a)).

	5 targets				**10 targets**			
#Fr	CC	SH	H	T	CC	SH	H	T
1	45.1	45.7	31.3	27.4	30.8	30.1	17.4	15.3
3	46.0	43.7	29.9	27.7	30.8	28.9	16.4	15.3
5	46.3	44.3	29.6	27.40	31.4	28.0	16.7	15.8
10	47.6	45.4	30.7	29.8	31.3	31.2	17.7	15.5
15	47.5	47.9	31.9	30.6	31.5	32.0	20.3	16.9
20	49.3	48.1	34.0	32.0	30.9	33.6	21.5	18.4
40	48.7	49.5	36.0	32.8	33.5	32.5	21.0	16.8

TABLE I: Synthesised recognition rates (%) from Fig 10 for 5 and 10 targets with increasing number of frames. The best #frames is shaded for each model. [Models: CC = culture colour, SH = soft histogram, H = height, T = texture]



(a) Culture Colours (b) Soft Histogram

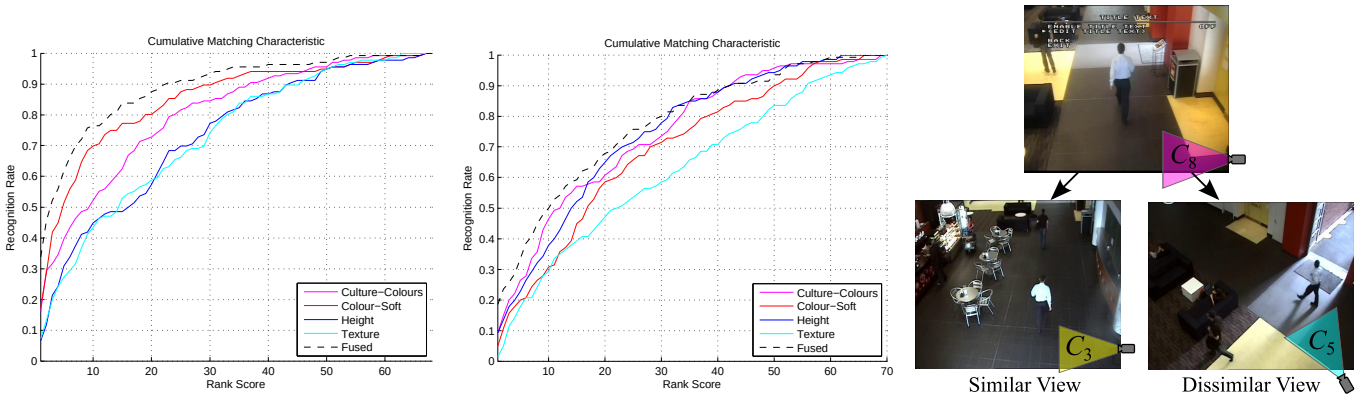
(c) Height (d) Texture

Fig. 10: Effect of number of frames used in the model when building models from a single camera view. All camera views are considered in this evaluation, with gallery and probe models trained off separate views. (See Table I for values at 5 and 10 targets)

B. *Effect of viewing angle*

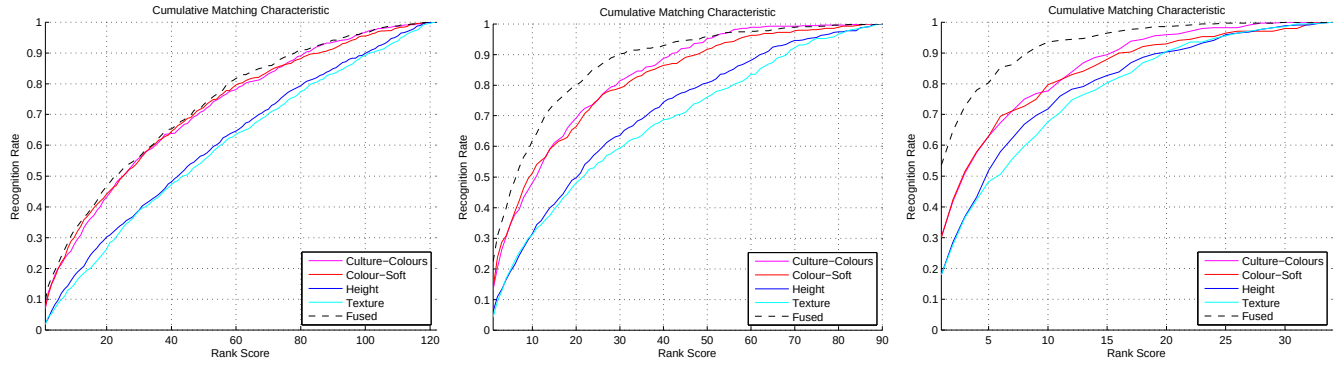
To evaluate the effect of viewing angle, we limit evaluation of testing and training models to two camera views which are similar in the captured viewing angle of a subject (Camera 3 and 8), and two camera views which are dissimilar (Camera 5 and 8). We make the assumption that subjects generally walk straight through the building and do not turn around, which holds true for the majority of subjects. It can be seen in Figure 3, that if this assumption holds (e.g. subjects walk left to right or right to left in the building diagram), we will obtain similar subject viewing angles in Camera 3 and 8 and dissimilar angles in Camera 5 and 8 as in Figure 11 (c). Results are presented in Figure 11.

It can be seen that all models degrade in performance with dissimilar views (recognition rates in Figure 11 (b) are lower than (a)), except for height which works similarly in differing viewing conditions (e.g. Height Rank-10 performance only degrades slightly, from 45% to 38%, while Colour-Soft degrades significantly from 70% to 31%), suggesting that height is more view invariant. This is expected, as height does not change from different viewing angles while colour and texture of a person may be different from the front/side/back. The full soft colour model outperforms culture colours in similar viewing angles (Figure 11 (a)), but culture colours perform better than full colour in differing viewing angles (Figure 11 (b)) and generally better across all camera conditions (Figure 12), suggesting that culture colours or other heavily quantised learned colour spaces are more stable than full colour in varied viewing conditions. The degradation in performance in the



(a) Similar view (Cam 3 and 8) (b) Dissimilar view (Cam 5 and 8) (c)

Fig. 11: The effect of viewing angle mismatches in training and testing. Evaluations consider gallery and probe models trained on separate views, with models built off 20 images. (a) shows CMC plots where testing and training models contain similar viewing angles, while in (b) testing and training models are built from dissimilar viewing angles. (c) displays example frames of a person in the selected similar and dissimilar camera views.



(a) 1 View (b) 2 Views (c) 3 Views

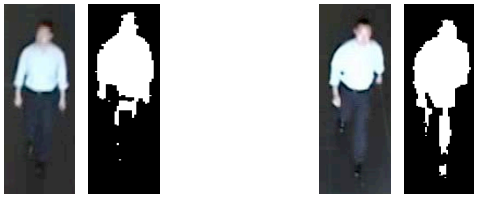
Fig. 12: CMC plots for colour, size, texture models, trained and tested on 1, 2 and 3 camera views using 20 images each.

colour and texture models may be attributed to the fact that many of the subjects appear different from the front, side and back due to items they are carrying (backpacks, shoulder bags) and their clothing (such as open jackets). However, considering all viewing angles, as in Figure 12, it can be seen that colour features are more discriminative and robust to all variations.

C. Effect of the number of viewpoints

In Figure 12, plots are presented for models trained on 1, 2 and 3 views. We consider all cameras and use 20 frames, with mutually exclusive views used in gallery and probe models. Colour models consistently outperform the height and texture models, and all models improve as more views are used to train the models. The improvement as more views are used is expected, as more information is included in the model. By including different viewing angles, the models better represent the person's overall appearance.

The superior performance of the colour models compared to height is expected, as there is more variation in colour, as heights will only differ by a few centimetres between subjects. Also, the height model is more affected by errors in segmentation (both of foreground pixels and segmentation into head, torso and legs). Small errors in the silhouette can result in a difference of a few centimetres or more, depending on where in the image the subject appears. While the colour biometric is also susceptible to segmentation errors, the colour models are less affected, except where segmentation errors result in large portions of the person not being visible (e.g. their legs or torso are not detected, as in Figure 13 (a)), or a large portion of the background being included in the model. The poor performance of the texture models may be caused by poor resolution which results in blurring of texture, and the lack of textural information in the majority of subjects. However, texture performs fairly consistently in differing conditions. In all cases, a fused model outperforms all individual models, as the complementary information from each model combined gives greater discrimination between people.



(a) Poor segmentation (b) Better segmentation

Fig. 13: An example of (a) poor segmentation and (b) better segmentation. Poor segmentation can result in missing body parts and reduce performance of the models. With many frames available, frame selection criteria can be used to filter out poorly segmented frames.

VI. Conclusion

In this paper we have presented a new database for the evaluation of person re-identification models in real surveillance conditions. Using the baseline models, we have shown how this new database can be used to better evaluate person recognition models in variable real-world conditions. In particular, we have demonstrated how this dataset can be used to evaluate a number of scenarios related to number of frames, number of cameras and viewing angles which can only be evaluated with a database consisting of a large number of subjects in a variable and unconstrained environment.

With the baseline models, it was found that colour models perform better across all viewing angles as there is greater discrimination in the models compared to height and texture. However, when considering exclusively different viewing angles, height was found to be quite stable, with colour and texture seen to be more view specific, as many subjects in the dataset appear different from the front, side and back due to carrying of objects (e.g. backpacks) and clothing characteristics (e.g. open jacket). It was also observed that culture colours (a quantised set of 11 colours) are slightly more stable than full colour histograms, suggesting that a heavily quantised learned colour space is preferable when encountering view mismatch.

In future work, methods to better fuse models with knowledge of the acquisition conditions will be explored to take advantage of the qualities of each model.

Acknowledgment

This research was supported by the Queensland Government's Department of Employment, Economic Development and Innovation, and the Australian Research Council's Linkage Project "Airports of the Future" (LP0990135).

References

[1] A. K. Jain, S. C. Dass, and K. Nandakumar, "Soft biometric traits for personal recognition systems," *Lecture notes in computer science*, 2004.

[2] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," *European Conference on Computer Vision (ECCV)*, 2008.

[3] A. Ess, B. Leibe, and L. Van Gool, "Depth and appearance for mobile scene analysis," in *International Conference on Computer Vision (ICCV)*, 2007.

[4] J. M. Ferryman, Ed., *Proc. of PETS2006*.

[5] S. Yu, D. Tan, and T. Tan, "A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition," in *International Conference on Pattern Recognition (ICPR)*, 2006.

[6] U. H. Office, "Imagery library for intelligent detection systems (i-LIDS) multiple camera tracking scenario definition," 2008. [Online]. Available: www.homeoffice.gov.uk/science-research/hosdb/i-lids/

[7] M. Hu, W. Hu, and T. Tan, "Tracking people through occlusions," in *International Conference on Pattern Recognition (ICPR)*, 2004.

[8] S. Denman, C. Fookes, A. Bialkowski, and S. Sridharan, "Soft-Biometrics: unconstrained authentication in a surveillance environment," *Digital Image Computing: Techniques and Applications (DICTA)*, 2009.

[9] S. Bak, E. Corvee, F. Bremond, and M. Thonnat, "Multiple-shot human re-identification by mean riemannian covariance grid," in *Advanced Video and Signal-Based Surveillance (AVSS)*, 2011.

[10] F. Tang, S. Lim, and N. Chang, "An improved local feature descriptor via soft binning," in *International Conference on Image Processing (ICIP)*, 2010.

[11] C. Madden, M. Piccardi, and S. Zuffi, "Comparison of techniques for mitigating the effects of illumination variations on the appearance of human targets," in *Advances in Visual Computing*. Springer, 2007.

[12] O. Javed, K. Shafique, and M. Shah, "Appearance modeling for tracking in multiple non-overlapping cameras," in *Computer Vision and Pattern Recognition (CVPR)*, 2005.

[13] G. Wu, A. Rahimi, E. Chang, K. Goh, T. Tsai, A. Jain, and Y. Wang, "Identifying color in motion in video sensors," in *Computer Vision and Pattern Recognition (CVPR)*, 2006.

[14] O. Hamdoun, F. Moutarde, B. Stanciulescu, and B. Steux, "Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences," in *International Conference on Distributed Smart Cameras (ICDSC)*, 2008.

[15] N. Gheissari, T. B. Sebastian, and R. Hartley, "Person reidentification using spatiotemporal appearance," in *Computer Vision and Pattern Recognition (CVPR)*, 2006.

[16] L. Bazzani, M. Cristani, A. Perina, M. Farenzena, and V. Murino, "Multiple-shot person re-identification by hpe signature," in *International Conference on Pattern Recognition (ICPR)*, 2010.

[17] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *Computer Vision and Pattern Recognition (CVPR)*, 2010.

[18] P. Forssén, "Maximally stable colour regions for recognition and matching," in *Computer Vision and Pattern Recognition (CVPR)*, 2007.

[19] S. Bąk, E. Corvee, F. Brémond, and M. Thonnat, "Person re-identification using Haar-based and DCD-based signature," in *Advanced Video and Signal Based Surveillance (AVSS)*, 2010.

[20] W. Schwartz, A. Kembhavi, D. Harwood, and L. Davis, "Human detection using partial least squares analysis," in *International Conference on Computer Vision (ICCV)*, 2009.

[21] H. Bouma, S. Borsboom, R. J. M. den Hollander, S. H. Landsmeer, and M. Worring, "Re-identification of persons in multi-camera surveillance under varying viewpoints and illumination," in *Proc. SPIE*, vol. 8359, 2012.

[22] M. Hirzer, C. Beleznai, P. Roth, and H. Bischof, "Person re-identification by descriptive and discriminative classification," *Image Analysis*, 2011.

[23] B. Prosser, W. Zheng, S. Gong, T. Xiang, and Q. Mary, "Person re-identification by support vector ranking," *British Machine Vision Conference (BMVC)*, 2010.

[24] R. Y. Tsai, "An efficient and accurate camera calibration technique for 3D machine vision," in *Computer Vision and Pattern Recognition (CVPR)*, 1986.

[25] S. Denman, C. Fookes, and S. Sridharan, "Improved simultaneous computation of motion detection and optical flow for object tracking," in *Digital Image Computing: Techniques and Applications (DICTA)*, 2009.

[26] S. Denman, A. Bialkowski, C. Fookes, and S. Sridharan, "Determining operational measures from multi-camera surveillance systems using soft biometrics," in *Advanced Video and Signal-Based Surveillance (AVSS)*, 2011.

[27] ——, "Identifying customer behaviour and dwell time using soft biometrics," *Video Analytics for Business Intelligence*, 2012.

[28] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence (PAMI)*, 2002.