Deep Learning Enabled Open World Thermal Re-Identification

Student Name: T.A. Robson

Supervisor Name: T.P. Breckon

Submitted as part of the degree of MEng Computer Science to the

Board of Examiners in the School of Engineering and Computing Sciences, Durham University

***Abstract* — Context/Background**

Although use of thermal imagery currently poses significant advantages for 24/7 surveillance in terms of the visibility of targets under all environmental conditions, a key limitation is the lack of colour information. Person re-identification across multiple cameras is a key research problem within the domain of visual surveillance and a important challenge for the future deployment of thermal sensing as an autonomous sensor technology. Many current approaches to the problem rely on colour features [8]. We have previously attempted to use a set of similar features to solve the thermal re-identification problem with little success [18], and are now exploring alternatives.

Aims

The aim of this project is to develop a system that would build upon and extend the range of existing thermal image detection, tracking and classification approaches. We want to build a system that can detect a person within thermal imagery in real time, distinguish a person from other objects and track a person moving through a scene in real time. We must then be able to re-identify people based on those our system has already seen. It will be trained on a set dataset, and then tested on a set of different people. This means that we are aiming to solve the open world re-identification problem, rather than simply to train and test in a closed world on the same set of people, as many previous approaches have [24].

Method

The first step for this is to be able to detect a person within thermal imagery in real time, distinguish a person from other objects and track a person moving through a scene in real time. This is done using a combination of a Mixture of Gaussians (MoG) background subtractor, a Histogram of Oriented Gradients (HOG) person detector and a Track-Learn-Detect (TLD) tracker. We then use a siamese Convolutional Neural Network to determine if a pair of images contains the same person or two different people. This architecture has been chosen as it enables us to deploy our solution on any camera system with any human targets, enabling us to achieve our aim of solving the open world re-identification problem.

Results

Conclusions

Keywords: Deep Learning, Siamese Network, Open World, Computer Vision, Person Re-Identification, Re-ID, Person Tracking, Track-Learn-Detect, Thermal Imagery, Thermal Video

I INTRODUCTION

A fundamental task for a distributed multi-camera surveillance system is to associate people across different camera views at different locations and times [8]. This is referred to as the person re-identification problem [8] and is an interesting and important problem within the field of computer vision. From the previous work, we can see that a substantial amount of research that has been done on person re-identification, mainly revolving around the use of features or attributes of a person [11]. However, much of this relies on the visible spectrum, with attributes of the form “red shirt” [11]. However, in the modern world, thermal imagery is often used for 24/7 surveillance when varying environmental conditions are present. Therefore, it is important that an effective re-identification system is developed to utilise this area of surveillance, as this problem has yet to be effectively solved.

There are many potential applications for this technology, but the most important in the modern world would be to support human intelligence organisations. The surveillance data that a system like this can provide would be critical for crime-prevention, forensic analysis, and counter-terrorism activities in both civilian and governmental agencies alike. While this surveillance data is currently widely used by human operators, these operators have to be trained, which offsets the utility of this approach with training and staffing costs. The implementation of an automated re-identification system is therefore of great interest, as it would be very useful in supporting these human operators and enabling them to achieve better results more efficiently. Figure 1 shows five different views of the same person that our features must be able to re-identify as being the same person.

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

Figure 1: Several frames showing the same person that our system must re-identify as the same person

Whilst thermal imagery has many advantages, it is not able to identify colour, making features that rely on colour useless. Therefore, alternative features are required to facilitate re-identification. In our previous work in [18], we attempted to use features that a human would think were distinctive enough to facilitate re-identification. This had only limited success. Since our work in [18], the attention of the researchers of re-identification has shifted to deep learning based solutions, as shall be discussed in more detail in the Related Work section. These deep learning approaches can be split into two categories: closed world, where the system is trained and tested on the same set of people from a pre defined dataset, and open world, where the system is trained to learn what makes people different, so it can be applied to any dataset of people. Figure 2 shows the difference between these two possible approaches to deep learning enabled re-identification.

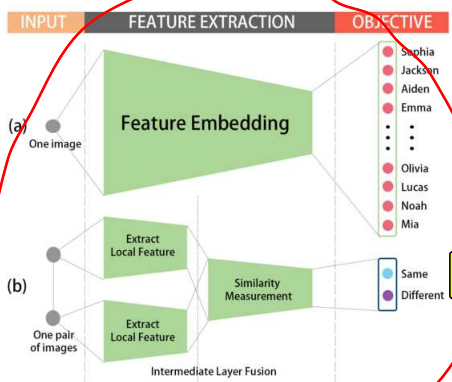


Figure 2: The difference between closed world (a) and open world (b) re-identification systems, from [24]

In order to improve the performance and runtime of our new approach, we are replacing the Kalman filter used in [18] with an implementation of the Track-Learn-Detect (TLD) tracker, originally proposed by the authors of [10]. The aim here is to ensure that we do not have to pass every frame to the neural network, but instead, when a person has been identified, they can then be tracked for as long as they remain unobscured in the frame, and continue to be labelled as the same person.

A Aims of the Project

The intent of this project is to develop a system that can utilise an open world deep neural network, paired with a real time person detection system and TLD tracker, to solve the thermal re-identification problem, and improve on the performance of state of the art solutions, including our own previous work [18].

B Achievements of the Project

II RELATED WORK

A Feature-Based Attempts at Re-Identification

Re-identification in colour is a well researched area, particularly using distinctive features for this purpose. Many of these methods are discussed in [8], and have been widely researched and understood, so research has moved on to explore more complex methods. An important part of the current state of the art in this area is camera network layout and topology, as explored in [14]. Here, the technique of distance vector routing is employed to get an idea of the relative locations of the cameras, enabling the system to prioritise the people seen most recently by the closest camera, as these are most likely to be correct. This is done by first analysing the overlap between cameras, and then computing distance vectors and probabilities of going from one camera to another, reducing the time complexity of the re-identification process in the majority of cases.

The work in [5] is on a similar theme to [14], but assumes a non-overlapping camera system. Each camera has entry and exit zones from its field of view, and if a person can get from one camera field of view to another they are directly connected. The system can then create what is referred to as a camera link model, using a temporal, spatial and appearance relation between the entry and exit zones of the cameras. These paths are obtained from training data, but the system itself learns how to recognise people by attributes, and uses the training data to estimate where they are most likely to have gone after leaving a given camera field of view.

The authors of [21] propose a different method for feature based identification, using a feature projection matrix to project image features of one camera to the feature space of another, to effectively eliminate the difference of feature distributions between the two cameras. This feature projection matrix is obtained through supervised learning. The proposed method aims to use a simple gradient descent algorithm to accelerate and optimise the re-identification process by compensating for the inconsistency of feature distributions captured by different cameras.

The work in [22] emphasises the importance of making good use of all images and video frames captured of a target. The system proposed here creates a gallery of images of known individuals, with more images increasing the accuracy of the system. When a gallery exists for a target, this is referred to as multi-shot re-identification, and single-shot re-identification when only one image is available in both the query and the gallery. For multi-shot re-identification, the authors propose to use geometric distance in another way by collaboratively approximating the query sets using all galleries, a method known as Collaborative Sparse Approximation.

Another approach, taken by the authors of [12], relies not only on extracting a set of features to use to compare people, but also on determining which of these features is the most discriminative for each person individually on the fly. This is achieved through the use of a random forest classifier, and functions well in an open world setting. However, since the time of publication of this work, the state of the art of this area has moved on to deep learning.

Our own previous attempt at thermal re-identification in [18] attempted to use features that a human would consider useful to re-identify people, such as an approximation of the shape of the target, an analysis of their gait and a measure of where the thermal hotspots of each person were. This performed resonably well for some features, but its inaccuracy in many situations led us to explore an alternative approach, using deep learning, based on the current state of the art in the area.

B Use of Deep Learning for Re-Identification

The use of deep learning to facilitate re-identification has been the driving force in the research community in this area in recent times. As our ability to train deeper and deeper networks on larger and larger datasets has increased, largely thanks to modern improvements in graphics hardware, these approaches become more and more relevant and powerful. As before, much of this research is focused on the colour spectrum. The work of [15] shows us how a combination of deep learning and human recognisable features can be used for re-identification. This works by training a neural network to recognise certain features that the authors consider to be discriminative. The result from the network is an ordered vector for the presence or absence of these features. However, as our previous work showed that we were unable to choose suitably descriminitive features for thermal re-identification, we will not be able to follow this approach.

Much of the rest of the previous deep learning work is split into open world and closed world. The work of [24] presents a useful comparison between these two problems, as shown in Figure 2. They refer to the closed world problem as Identification and the open world problem as Verification, with the justification that in closed world, the aim is to identify which of a given set of people the target is, where as in open world, the aim is to verify whether two targets are the same or not. These two approaches differ greatly in terms of input method, feature extraction and the loss function used. The authors of [24] conclude, as we have, that the open world problem is more relevant to real world applications.

A solution to the open world problem is attempted in [23]. The authors aim to combine feature learning and re-identification into one framework, which is a deep siamese CNN with an SVM placed on top of the network, after the last layer, to attempt open world re-identification. The results of this are promising for the potential use of such an architecture, but fall short of the state of the art at time of writing. The apparent reason for this is overfitting to the dataset used to train the architecture, and the authors hypothesise that if they could use a dataset with wider variation of people, improved results could be achieved. As this work is several years old, we can hypothesise that with modern graphics hardware, this model could realistically be trained on a much larger and more varied dataset, which should improve the results. This hypothesis is backed up by the authors of [1], who have implemented architecture and trained on a larger dataset which consistently outperforms the state of the art at the time of its writing.

The work in [20] proposes a similar deep siamese CNN architecture, but also retains the single image representation. The purpose of this is to reduce the amount of computational work that must be done on both images together, as well to joint optimise the single image representation and cross image representation for improved accuracy at a lower computational cost. The authors claim to outperform the majority of the state of the art methods at the time of writing. The work presented in [6] uses a pair of siamese CNNs, one to learn spatial information and the other to learn temporal information. These features are then weighted, as the authors claim that spatial features are more discriminative than temporal features. This method outperforms or shows comparable results to the existing best performing methods. The work in [17] is another siamese CNN with the particular aim of making it work for people represented at different scales. This would be an interesting addition to our work, but is currently out of scope for us.

The authors of [2] do not propose a novel architecture as such, but instead propose a set of good practices that should be followed for effective re-identification, such as pre-training for identity classification, sufficiently large image resolution, state of the-art base architecture and dataset augmentation with difficult examples. We will endeavour to follow these guidelines in our work in this project wherever possible.

C TLD Tracker

The Track-Learn-Detect (TLD) tracker was first proposed by ~~Zdenek~~ Kalal in [10]. The idea of this tracker is to break down the person-tracking task into tracking, learning and detection. The tracker follows the object between frames. The detector corrects the tracker if necessary based on previous observations. The learning estimates detector's errors and updates it to avoid these errors in the future. However, the main purpose of TLD is to be able to consistently follow the same object through as changing background, for example a car driving along a road being recorded from a helicopter. The authors have said that it performs sub optimally on varying targets such as people. During our early experimentation, we found that is was sufficiently capable of tracking a person through a single camera frame, but could not re-identify people across multiple cameras, or even returning from behind an obstruction.

D Relation to this Project

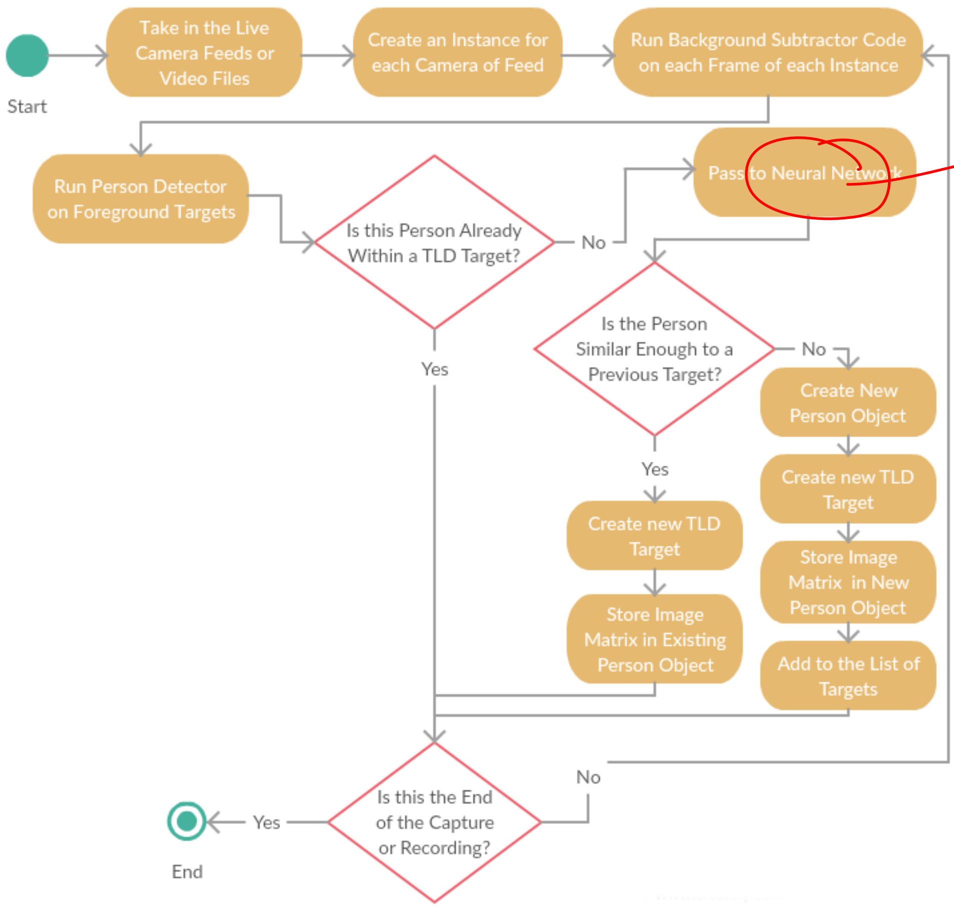
In this project we use elements from many of the papers discussed in this section. The TLD tracker will be used simply for tracking a person ~~for as long as they~~ remain visible in one camera view. Whether this person has been previously observed by the system or not will be determined by an open world siamese CNN, with an architecture similar to the related work. We take inspiration from [20] and will extract the features of each individual image before comparison between them. However, many of these papers that propose such a network only show results when it is trained and tested on a specific dataset of still images. They do not integrate it with a tracking system to see how it can perform in the real world. Also, this previous work has all been done on colour imagery. Very little work has been done to try and solve the re-identification problem in thermal imagery, save for our own previous work [18]. Therefore, the major research aim of our project, and where we are advancing the state of the art, is to see if such a network architecture can successfully be applied to thermal imagery, and whether it can perform effectively as a part of a full re-identification system in real time.

III SOLUTION

Having established the problem that we want to solve, we will now break down the most important elements of our solution, giving an overview of the structure of our implementation and a description of the elements used.

A Implementation Structure and Tools Used

The implementation will begin by opening each video file, or live camera feed, and concurrently running the real time target detection code on these. Each time this code identifies a person that does not currently have a TLD tracker object associated with them, it compares this person to the other people that have been seen previously using the siamese CNN, and if they are deemed to be sufficiently similar to one of these people, then they are re-identified as the same person, else the system creates a new person object. Each of these person objects has an associated TLD tracker and set of previous observations, and these are used to facilitate the comparison between targets, and are updated each time the target is successfully identified. This continues frame by frame until the video file or camera feed ends, as shown in Figure 3.



CNN

Figure 3: Logic of our Re-Identification System

Many of the computer vision techniques that will be described in the next section are complex to implement. Therefore, ~~version 3 of the~~ OpenCV library [3] has been used to allow us to use stable, well tested code. However, the implementation of TLD code in OpenCV is suboptimal, so we have used the existing open source multi-object implementation of TLD from [13]. The neural network side of the project is written in Keras [4], using the Tensorflow backend.

B Real Time Person Detection

Before we can start concerning ourselves with deep learning based person re-identification, we must be able to identify whether a person is present in the image at all, so that we have a region of interest to pass to the neural network, and so we can use TLD to track targets through single camera views.

B.1 Background Subtraction

The first stage of this process is background subtraction from the static camera viewpoint. We use the Mixture of Gaussians (MoG) technique to facilitate this, taking inspiration from [25, 26]. This technique works by building up a background model over multiple camera frames, modelling each of the pixels using multiple Gaussians. Using a Gaussian over the last N frames, where N is given by a parameter specifying the rate at which the background model is updated, is far more memory efficient than storing all the pixel values across the entire video capture. This update rate is determined by the trade off between being fast enough to absorb objects that have become stationary into the background and being slow enough to allow the detection of slow moving objects.

As the program runs through the video, during each new frame, a Gaussian for each pixel is evaluated using a simple heuristic to determine which is most likely to correspond to the background model. Pixels that do not match closely enough with these background Gaussians are classified as foreground elements and added to a new image in the code. Once these foreground pixels are identified and built into a foreground mask, erosion and dilation image operators [19] are used to clean up these results. From here, we use the contour detection to find the connected components and draw bounding boxes around these contours.

B.2 Person Identification

Having identified the foreground objects, we must now determine whether they are people. The implementation uses Histogram of Oriented Gradients (HOG), discussed in [7]. This method works by performing edge gradient calculation on the bounding box identified by the background subtractor. From here cell histograms are computed, with each of the histogram entries filled by gradient magnitudes. These histograms are then used to create overlapping block histograms of the adjacent cells. These block histograms are then concatenated to give a HOG descriptor, a high dimensional vector. This HOG descriptor is then passed to a pre-trained machine learning algorithm, in this case a Support Vector Machine (SVM). If this comes up with a positive identification, then it is classified as a person and will be given an associated TLD tracker object.

B.3 Person Tracking using Track-Learn-Detect

Once we know where the people are in the camera view, we can track their movement. We use the Track-Learn-Detect algorithm, originally proposed in [10]. This algorithm was originally intended to facilitate the long term tracking of an unknown object in a video stream. The algorithm works by breaking down the person tracking task into tracking, learning and detection. The tracker follows the object between frames. The detector corrects the tracker if necessary based on previous observations. The learning estimates detectors errors and updates it to avoid these errors in the future. In our early experimentation, we found that the learning part of this algorithm meant that it was attempting to perform re-identification when people left and re-entered a video stream, or appeared in a different one. Therefore, we had to cut some of the functionality out of TLD for this implementation, so that each object was deleted when it left the scene or became obstructed. This eliminated the miss-classifications that the limited re-identification capabilities of this algorithm were causing, enabling us to effectively track people across the scene, whilst using our neural network for the re-identification.

Therefore, in this implementation, each person currently present in any of the camera views has an associated TLD tracker object, which is created on the first HOG identification in this neighbourhood, and classified by the neural network. The tracker then follows this person through the frame until either the person leaves the frame or HOG is no longer getting a positive identification within the tracker. If either of these conditions occurs, the TLD object is deleted. In the case of no HOG identification being present, this indicated that the tracker has incorrectly predicted the position or rate of movement of the person and has lost them. A new tracker will then be created to replace it in the next frame when HOG detects this person again.

C Deep Learning

We can now move on to our approach to deep learning based re-identification, considering our dataset and our architecture choices.

C.1 Training and Testing Dataset

The model was trained on the dataset collected for our previous work [18], as this contained video of multiple people, some similar and some greatly differing, from different views. The real time person detection system discussed previously was adapted into a data extraction system, where the identification of each person was performed by a key press, rather than the neural network. Each successive frame of each person was then saved to a file, labelled with which person it showed. The size of these files was fixed at 258x128, to ensure that the size of the file had no effect on the training. This meant that the system was trained on varying resolutions, as some regions of interest had to be enlarged, and others had to be shrunk. This would help the ability of the system to re-identify at different distances away from the cameras.

These files were then formed into positive and negative pairs of images, where positive pairs were both images of the same person and negative pairs were not. We took care to ensure we had an equal number of positive and negative pairs, giving a balanced dataset. These pairs were assigned binary labels, 1 for positive pairs and 0 for negative pairs. We then applied some data augmentation to the images, following the advice of [2]. This involved applying a random transformation to each image, effectively doubling the size of the dataset. The possible transformations consisted of a horizontal flip, small shifts vertically or horizontally, and a small rotation. The aim of this is to introduce more variation into our dataset and enable it to cope better with difficult real world situations.

C.2 Network Architecture

Taking inspiration from the state of the art literature in the area of open world person re-identification, our network architecture is a Deep Siamese CNN. This means that the network is trained on pairs of images, and attempts to determine whether these images show the same person or a different person. The two networks which are each fed an image have exactly the same architecture and weights, as they are trained together. Our CNN consists of convolutional layers, pooling layers and fully connected layers.

The convolutional layers operate directly on the image to reduce the complexity of the input in a manner that is both meaningful and structured. This enables the network to have fewer neurons than a traditional fully connected feedforward network would have if operating on an image. This enables the convolutional neural network to be far deeper with fewer parameters. This resolves the vanishing or exploding gradients problem that would otherwise occur, as a greater number of training parameters in the early layers of the network would make it unstable, and prone to the weights becoming much too large (exploding gradients) or much too small (vanishing gradients).

Pooling layers then down-sample their input, looking at a neighbourhood of pixels and, in the case of max pooling that we have used here, output the maximum. The aim of this is to provide robustness to the changes in the spatial location of features across the dataset and to reduce input dimensionality. This is also an effective tool to control overfitting, as it reduces the number of training parameters and ensures that it is the general regions where features are present that are learned, not the specific pixels of the images in the training set in which they are present.

Finally, we flatten the output of our last pooling layer to a single dimension and then pass it to a fully connected layer. The high dimensional vector output of this layer will be the representation of this image, and we will use the euclidean distances between these vectors to determine whether the inputs show the same person or different people.

At various stages of the network, we have some dropout layers. The purpose of dropout is to eliminate a proportion of the training data at each epoch to help to prevent overfitting, as the network is being trained on a different dataset each epoch. The goal is therefore to force the network to learn more robust features that are useful in conjunction with random subsets of the other neurons so that good performance is still achieved when some neurons are removed.

The loss function to be used for this network was proposed by the authors of [9]. This paper defines a contrastive loss function, which maps high dimensional inputs to lower dimensional outputs, given distances between samples in its input space. These distances between samples are supplied by the euclidean distance. The formula for this function is given by equation (1) below, where Y is the label (0 or 1), m is the margin that determines the the maximum euclidean distance between two points that will influence the loss function. The purpose of this is to ensure that vastly different images do not have a disproportional effect on the learning outcome. D is the euclidean distance, which is defined here by equation (2) below, where X1 and X2 are the input images and F represents the feature vectors output by the siamese pair of identical CNNs.

$$L = (1-Y)\frac{1}{2}(D)^2 + Y\frac{1}{2}\{max(0, m-D)\}^2 \tag{1}$$

$$D = \sqrt{\{F(X1) - F(X2)\}^2} \tag{2}$$

The actual architecture chosen was two convolutional layers, a max pooling layer, a dropout layer, two convolutional layers, a max pooling layer, a dropout layer, a flatten layer, a dropout layer and a dense fully connected layer. The parameters for these layers such as the activation functions, the convolutional kernel sizes, stride sizes, pooling width, dropout rates and size of the dense layer were all determined by an exhaustive grid search using the Hyperas library [16].

FIGURE HERE FROM TENSORBOARD ONCE FINALISED

IV RESULTS

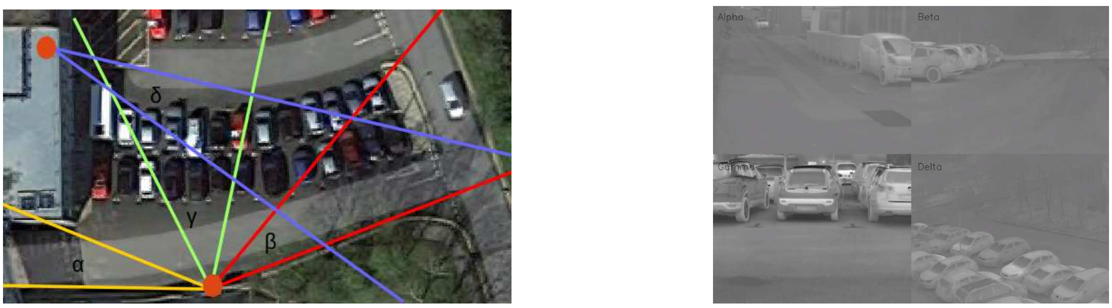
A *The Deep Learning System*

Train/test results, tensorboard grpahs, TNSE plots, excel plots of distances.

B *The Re-Identification System Dataset*

The data used to evaluate the re-identification system was gathered with three cameras at Durham University. The cameras and their fields of view are arranged as they are in figure 4(a), with the cameras labelled as camera α, β and γ respectively, and their fields of view shown by the matching coloured lines. The images seen by the cameras are shown in figure 4(b). The dataset contains five people.

CHANGE THIS FIGURE WHEN ALL RUNS CORRECTLY



(a) (b)

Figure 4: Position and field of view of the cameras used to record our data

C *Re-Identification Performance*

V EVALUATION

We will now evaluate the strengths, weaknesses and limitations of the research that has been presented here.

A Real Time Person Detection System

B Deep Learning System

C Re-Identification System

D Appraisal of Project Organisation

VI CONCLUSION

References

[1] Ejaz Ahmed, Michael Jones, and Tim K Marks. An Improved Deep Learning Architecture for Person Re-Identification. *Computer Vision and Pattern Recognition (CVPR)*, pages 3908–3916, 2015.

[2] Jon Almazan, Bojana Gajic, Naila Murray, and Diane Larlus. Re-ID done right: towards good practices for person re-identification. *CoRR*, 1801.0, 2018.

[3] G Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[4] François Chollet. Keras, 2015.

[5] Chun-Te Chu and Jenq-Neng Hwang. Fully Unsupervised Learning of Camera Link Models for Tracking Humans Across Nonoverlapping Cameras. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(6):979–994, jun 2014.

[6] Dahjung Chung, Khalid Tahboub, and Edward J Delp. A Two Stream Siamese Convolutional Neural Network for Person Re-identification. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 1992–2000, 2017.

[7] Navneet Dalal and Bill Triggs. Histograms of Oriented Gradients for Human Detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005.*, 2005.

[8] Shaogang Gong, Marco Cristani, Chen Change Loy, and Timothy M Hospedales. The Re-Identification Challenge. In *Person Re-Identification*, pages 1–21. 2014.

[9] Raia Hadsell, Sumit Chopra, and Yann LeCun. Dimensionality reduction by learning an invariant mapping. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 1735–1742, 2006.

[10] Zdenek Kalal, Krystian Mikolajczyk, and Jiri Matas. Tracking-Learning-Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(1), 2010.

[11] Ryan Layne, Timothy M Hospedales, and Shaogang Gong. Attributes-based Re-Identification. In *Person Re-Identification*, pages 93–121. 2014.

[12] Chunxiao Liu, Shaogang Gong, and Chen Change Loy. On-the-fly feature importance mining for person re-identification. *Pattern Recognition*, 47(4), 2014.

[13] Christian Lutz and Thorsten Engesser. MOTLD, 2015.

[14] Niki Martinel, Gian Luca Foresti, and Christian Micheloni. Person Reidentification in a Distributed Camera Network Framework. In *IEEE Transactions on Cybernetics*, pages 1–12. IEEE, 2016.

[15] Tetsu Matsukawa and Einoshin Suzuki. Person re-identification using CNN features learned from combination of attributes. In *Proceedings - International Conference on Pattern Recognition*, pages 2428–2433, 2017.

[16] Max Pumperla. Hyperas, 2017.

[17] Xuelin Qian, Yanwei Fu, Yu Gang Jiang, Tao Xiang, and Xiangyang Xue. Multi-scale Deep Learning Architectures for Person Re-identification. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2017-Octob, pages 5409–5418, 2017.

[18] Thomas A. Robson and Toby P. Breckon. *Camera-to-Camera Tracking for Person Re-identification within Thermal Imagery*. Bachelors thesis, Durham University, 2017.

[19] C J Solomon and T P Breckon. *Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab*. Wiley-Blackwell, 2010.

[20] Faqiang Wang, Wangmeng Zuo, Liang Lin, David Zhang, and Lei Zhang. Joint Learning of Single-Image and Cross-Image Representations for Person Re-identification. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1288–1296, 2016.

[21] Yimin Wang, Ruimin Hu, Chao Liang, Chunjie Zhang, and Qingming Leng. Camera compensation using a feature projection matrix for person reidentification. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(8):1350–1361, aug 2014.

[22] Yang Wu, Michihiko Minoh, Masayuki Mukunoki, Wei Li, and Shihong Lao. Collaborative sparse approximation for multiple-shot across-camera person re-identification. In *Proceedings - International Conference on Advanced Video and Signal-Based Surveillance*, pages 209–214. IEEE, sep 2012.

[23] Guanwen Zhang, Jien Kato, Yu Wang, and Kenji Mase. People re-identification using deep convolutional neural network. *VISAPP 2014 - Proceedings of the 9th International Conference on Computer Vision Theory and Applications*, 3:216–223, 2014.

[24] Zhedong Zheng, Liang Zheng, and Yi Yang. A Discriminatively Learned CNN Embedding for Person Re-identification. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 14(1), 2016.

[25] Zoran Zivkovic. Improved Adaptive Gaussian Mixture Model for Background Subtraction. *Proceedings of the 17th International Conference on Pattern Recognition*, 2004.

[26] Zoran Zivkovic and Ferdinand Van Der Heijden. Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern Recognition Letters*, 27(7):773–780, 2005.