# OPEN-SET PERSON RE-IDENTIFICATION THROUGH ERROR RESILIENT RECURRING GALLERY BUILDING

*Philine Witzig, Evgeniy Upenik and Touradj Ebrahimi*

Multimedia Signal Processing Group (MMSPG)
École Polytechnique Fédérale de Lausanne (EPFL)
CH-1015 Lausanne, Switzerland
Email: firstname.lastname@epfl.ch

## ABSTRACT

In person re-identification, individuals must be correctly identified in images that come from different cameras or are captured at different points in time. In the open-set case, the above needs be achieved for people who have not been previously recognised. In this paper, we propose a universal method for building a multi-shot gallery of observed reference identities recurrently online. We perform L2-norm descriptor matching for gallery retrieval using descriptors produced by a generic closed-set re-identification system. Multi-shot gallery is continuously updated by replacing outliers with newly matched descriptors. Outliers are detected using the Isolation Forest algorithm, thus ensuring that the gallery is resilient to erroneous assignments, leading to improved re-identification results in the open-set case.

***Index Terms*—** person re-identification, open-set person re-identification, image processing, deep learning

## 1. INTRODUCTION

Person re-identification (Re-ID) refers to the problem of associating a unique identity label (an identification) to a person appearing in an image or a video segment and retrieving the assigned label if the same person reappears in a new image or video segment that was captured by another camera or at another point in time.

Person Re-ID has great importance in various domains: be it video surveillance for security reasons at airports or finding a lost child in a theme park. An especially important domain becomes vivid in the context of the project in which this paper was carried out, namely, ProCam. ProCam aims at efficiently tracking individuals potentially infected by a transmissible disease such as COVID-19 through a smart camera setup for this purpose while at the same time protecting their privacy. Contact tracing of such individuals and analysis of their behaviour and interactions with others and their environment can be a useful tool to prevent the spread of contagious diseases. Here, we focus on the latter without addressing the challenge in protection of privacy that is out of the scope of this paper.

By applicability and scope, the Re-ID methods known today can be categorized in two main classes: closed-world and open-world settings, as summarized in a recent survey [1]. For the latter category a query person may or may not exist in the gallery. Thus, verification rather than retrieval is performed in order to discriminate whether two images represent the same person. Furthermore, for a specific type of open-world settings, that is open-set person Re-ID problems, the gallery is updated when new identities are encountered. A typical Re-ID architecture consists of the following main steps: 1) data acquisition, that may be performed by capturing images or video with a single camera or a set of cameras; 2) bounding box generation, that is usually performed by an external object detector; 3) descriptor creation, i.e. finding a vector in a feature space that discriminatively represents a person; 4) gallery creation, i.e keeping track of a set of descriptors of reference identities; and 5) person retrieval, i.e. associating new data with a stored identity. For the open-set case, step (4) is performed continuously and may start with an empty gallery. In particular, one descriptor per identity is kept for a single-shot gallery, while for a multi-shot gallery, multiple descriptors are stored for each identity.

Past research has focused mainly on solving the closed-world Re-ID problems through building better descriptors with the help of deep learning and suitable distance metrics. Less research, however, has been focused on the open-set Re-ID problem that is increasingly faced in public spaces.

We propose a simple method to solve the problem of gallery creation and maintenance in the open-set case. The method is universal and can be applied to extend a generic closed-world system. We build a multi-shot gallery online and continuously update the set of stored descriptors. Additionally, we make use of an outlier-based decision rule for inserting new identities. When matching a query descriptor to

This work was supported by MAKE, an interdisciplinary educational initiative at EPFL in the framework of ProCam project. Ph. Witzig performed the work while being an exchange student at EPFL.

an identity that is already available in the gallery, its descriptors are updated by performing outlier detection once more and replacing it by the matched query descriptor. This makes the gallery resilient to erroneous descriptor assignments.

## 2. RELATED WORK

Zheng et al. [2] were the first to propose an approach to the open-set Re-ID problem using a transfer ranking framework for set-based verification. Shortly after, Karaman et al. [3] proposed a solution based on conditional random field inference. With the rise of deep learning in the following years, the open-set problem moved into the background. Although discriminative descriptors (either appearance based [4, 5] or learned [6, 7, 8]) and reliable distance metrics [9, 10] are the keys for improving the accuracy of Re-ID systems, they are often evaluated in the closed-world settings.

Furthermore, while a few authors proposed solutions based on single-shot galleries [11], they lack the benefit of diverse appearance information about people kept in multi-shot galleries. Having different descriptors of each identity can make Re-ID systems more robust and more generic [12].

Vidanapathirana et al. [13] proposed a method to use already developed descriptors in the multi-shot open-set case. They perform novelty detection by keeping a relatively large set of random individuals. In situations where a query descriptor does not *strongly* match any descriptor in a random person set, they assume that the person in query must be a new identity. However, this approach is computationally expensive due to multiple comparisons within the random person set.

Furthermore, to the best of our knowledge, no approach exists that can be applied to a working closed-world system in order to extend it to multi-shot gallery building while being resilient to erroneous assignments.

## 3. UNDERLYING ARCHITECTURES

To obtain descriptors, we require a robust method optimized for a closed-world case that produces feature vectors that (i) are highly discriminative and (ii) can be interpreted in Euclidean space. Let $f_p$ be a descriptor produced by such a method for an image of person $p$ and $f_q$ a descriptor for an image of person $q$. Then we want the distance

$$d(f_p, f_q) = ||f_p - f_q||_2 \tag{1}$$

to be large, if $p$ and $q$ are different. On the other hand, L2 Euclidean distance should be small, if $p \equiv q$. After in-depth analysis of the state of the art , the AlignedReID network [14] drew our attention due to its high accuracy performance in closed-set person re-identification and a Euclidean descriptor space. The method produces local and global features, which are trained jointly. The authors claim that the global descriptors suffice for Re-ID and can be interpreted in Euclidean space. However, note that our method can be applied to any other similar (also non-neural) architecture that satisfies the criteria (i) and (ii).

## 4. GALLERY BUILDING METHOD

Let $f_q \in Q$ be a descriptor produced for a query image that has been cropped to the bounding box. Let $G = \{F_1, \ldots, F_n\}$ be a multi-shot gallery set. Initially, $G = \emptyset$. Furthermore,

$$F_i = \{f_{i,1} \ldots f_{i,m}\}, \forall i \in [1, n] \tag{2}$$

holds. $F_i$ corresponds to a collection of $m$ descriptors for identity $i$. We perform the identity assignment to $f_q$ as follows: for each $F_i \in G$, we first compute the mean distance to $f_q$ through

$$\overline{d}(F_i, f_q) = \frac{1}{m} \sum_{j=1}^{m} ||f_{i,j} - f_q||_2 \tag{3}$$

and store the $i^*$ for which $\overline{d}$ is minimal (Algorithm 1). This corresponds to a reference identity that is closest to the query out of the stored reference identities. We test if our getOutlier() function detects $f_q$ to be an outlier for $F_{i*} \cup f_q$. If this is the case, we assume that $f_q \in U$, where $U \subseteq Q$ is the set of descriptors of unknown identities. Thus, a new identity $n+1$ is created and the corresponding multi-shot set $F_{n+1} = \{f_q\}$ is added to $G$. If for getOutlier($F_{i*} \cup f_q$), $f_q$ is not an outlier, we check if we have reached the maximum number of shots per identity $m$ in $F_{i^*}$. If the maximum is reached, we replace one of the descriptors $f_{i^*,j}$ which corresponds to an outlier. Otherwise, $f_q$ is simply added to $F_{i*}$. This is illustrated in Figure 1.

**Algorithm 1:** Error Resilient Gallery Building

```
Data: G, f_q
for F_i in G do
    i* = arg min_i d̄(F_i, f_q);
    f_out = getOutlier(F_i* ∪ f_q) ;
    if f_out == f_q then
        F_{n+1} = {f_q};
        G = G ∪ {F_{n+1}}
    else
        if len(F_i*) == m then
            F_i* = F_i* \ getOutlier(F_i*);
        end
        F_i* = F_i* ∪ {f_q};
    end
end
```

Since the feature space is of high dimension, i.e. $\mathbb{R}^{2048\times1}$, an outlier detection method that performs well for high-dimensional data is required. Distance based methods are not

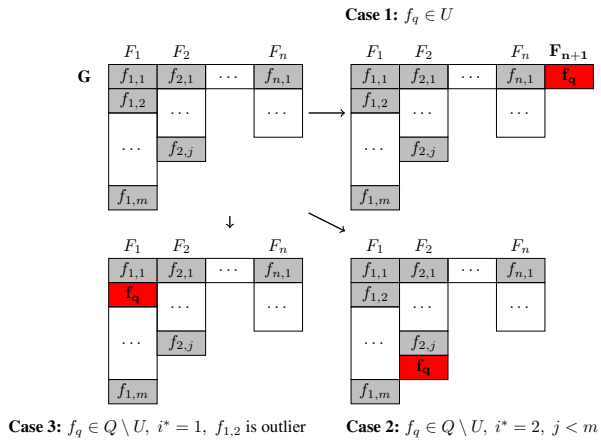



**Fig. 1**: Visualization of the three cases for updating the gallery $G$ (top left). **Case 1**: $q$ is detected to be a new identity (top right), **Case 2**: $q$ is detected to be a known identity and matched with $i^* = 2$, $|F_2| < m$ (bottom right). **Case 3**: $q$ is detected to be a known identity and matched with $i^* = 1$, $|F_1| = m$ and $f_{1,2}$ is detected to be an outlier (bottom left).

suitable in this context because the high-dimensional feature space is sparse. For this reason, we opted for the Isolation Forest algorithm [15], which aims at isolating anomalies. More precisely, each sample is assigned an anomaly score. This score is computed based on the path length in the isolation tree that the sample produces during random feature splitting, and anomalies produce shorter paths. In the case multiple outliers are detected using this algorithm, we arbitrarily select one outlying descriptor that is replaced by $f_q$. In the case no outlier is detected, we randomly replace one of the descriptors. With this approach, on the one hand, we make sure that wrong assignments are removed from the multi-shot gallery, thus making the gallery building error-resilient, and on the other hand, we always keep the most recent observations of an identity. For implementation details, please refer to our GitHub repository `https://github.com/mmspg/ERRGB`.

## 5. EXPERIMENTS AND RESULTS

### 5.1. Quantitative Analysis

Our approach is evaluated quantitatively on the evaluation set of the Market_1501 dataset. The test dataset is adapted for the open-set case by computing a random hold-out set $U \subseteq Q$. The identities associated with the descriptors contained in the hold-out set are removed from the gallery, i.e. $G' = G \setminus U$ since they correspond to the descriptors of unknown identities. Furthermore, we limit the number of shots per identity in the gallery to $m = 50$. During testing, $G'$ is constantly updated with matched identities.

To measure the performance, we use three metrics: an adapted version of rank-1 accuracy (suited for the open-set ReID problem), true target recognition (TTR) and false target recognition (FTR) [16]. TTR and FTR are two metrics that are particularly designed for the open-set case where "target" refers to the known identities. The three metrics are computed as follows:

$$rank_1 = \frac{1}{|Q|}\left( \sum_{q=1}^{|Q\setminus U|} r_1(q) \ + \ \sum_{q=1}^{|U|} \delta_{f_{out},f_q} \right), \quad (4)$$

where $r_1(q)$ is an indicator function which is 1 if $q$ is at rank 1 (and 0 otherwise) and $\delta_{f_{out},f_q}$ the Kronecker delta (i.e. $q$ was correctly identified as a new identity). Moreover,

$$TTR = \frac{r_1(q)}{|Q \setminus U|} \quad (5)$$

and

$$FTR = \frac{|U| - \sum_{q=1}^{|U|} \delta_{f_{out},f_q}}{|U|}. \quad (6)$$

The TTR divides the number of query images of known identities that were verified as one of the known people by the total number of query images of known identities. In contrast, FTR measures how many unknown identities were verified as one of the known people out of all query images corresponding to unknown identities. In Table 1, we report rank-1 accuracy, TTR and FTR for different sizes of the hold-out set $U$. Table 2 shows the performance of state-of-the-art person ReID methods that were evaluated on Market_1501 [17]. We compare our approach to the state of the art, where TTR is displayed against a predefined set of FTR values, as it is done in literature [17]. Since it is not possible in our experimental setup, to pre-define FTR values and obtain the corresponding TTR values for this predefined set, we approximately locate the corresponding TTR values in Table 2 using the results from Table 1. This means that according to Table 1, where for an FTR of $5.9\%$ the corresponding TTR is $85.7\%$, we assume that for an FTR of $5\%$ the corresponding TTR $\approx 85.7\%$. Likewise, the FTR of $10\%$ corresponds to TTR $\approx 83.5\%$.

**Table 1**: Results on Market_1501 modified for the open-set case with different sizes of $U$.

| $\lvert U\rvert$ | $rank_1$ | $TTR$ | $FTR$ |
|---|---|---|---|
| 100 (13.3%) | 84.8% | 83.5% | 9.1% |
| 375 (50.0%) | 89.6% | 84.4% | 6.2% |
| 500 (66.7%) | 91.7% | 85.7% | 5.9% |

### 5.2. Qualitative Analysis

Moreover, we evaluated the approach qualitatively on custom video data. The following attributes related to a recording were manipulated: number of people in front of the camera, distance to the camera, illumination of the scene and type of

**Table 2**: Comparison of open-set person ReID methods on Market_1501: TTR (%) against FTR.

| Method | 0.01% | 1% | 5% | 10% | 20% | 30% |
|---|---|---|---|---|---|---|
| ERRGB (Ours) | - | - | $\approx \mathbf{85.7}\%$ | $\approx \mathbf{83.5}\%$ | - | - |
| APN [17] | 9.01 | 22.32 | 46.78 | 63.34 | 73.82 | 81.12 |
| DCGAN+LSRO [18] | 6.77 | 20.60 | 42.06 | 58.80 | 72.49 | 80.11 |
| DeepFool [19] | 0.78 | 21.89 | 45.05 | 59.23 | 69.53 | 85.41 |

**Table 3**: Qualitative analysis manipulating different parameters related to the recording.

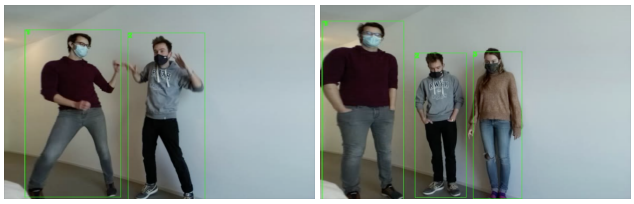
| # People | Dist. | Lighting | Actions | Stability |
|---|---|---|---|---|
| 2 | Close | Flat | Static | yes |
| **2** | **Close** | **Flat** | **Dynamic** | yes |
| 2 | Close | Illuminated | Static | no |
| 2 | Close | Illuminated | Dynamic | no |
| 2 | Far | Flat | Static | no |
| 2 | Far | Flat | Dynamic | no |
| **3** | **Close** | **Flat** | **Static** | yes |
| 3 | Close | Flat | Dynamic | no |

movement (cf. Table 3). The video sequences were recorded using the Raspberry Pi camera module V2 mounted to the Raspberry Pi 4 Model B. Person detection was performed using YoloV3 [20] and we used a correlation tracker. There were either two or three people in front of the camera with distances to the camera ranging from "Close" ($2.5 - 3$m) to "Far" ($5$m). The lighting conditions varied by either having strong background illumination (cf. "Illuminated"), or not (cf. "Flat"). Participants actions were either "Static", i.e. facing the camera frontally, or "Dynamic", i.e. moving around and crossing each other. Figure 2 shows two extracts from video footage corresponding to the trials in bold from Table 3 and the computed label predictions. Note that the gallery was initially empty. We evaluated each run by judging whether the label assignments would be stable or not. Unstable assignments would correspond to label swapping between participants (re-assignment of a label) or assigning multiple labels to one person.

## 6. DISCUSSION

According to the quantitative results achieved on the validation set of Market_1501 modified for the open-set case, our method significantly outperforms related work in terms of TTR/FTR metrics. Moreover, we achieve high rank-1 accuracy that is comparable to the closed-set rank-1 accuracy (cf. $94.4\%$ [14]). Finally, the performance tends to improve with the size of the hold-out dataset, i.e. the number of unknown identities. Our assumption is that it is easier in our method to decide, whether a query is an unknown person or not, if there are not many known people in the gallery it can choose from.

In terms of qualitative analysis, Table 3 indicates that our method works well in practice if we have a controlled environment. In particular, people must be close to the camera. This might be due to a rather small resolution of the Raspberry Pi camera module V2. Moreover, strong background illumination causes label assignments to become unstable. We assume that this is due to the data for which the underlying architecture (AlignedReID) was trained on, where we have constant ambient illumination without strong background illumination. Finally, we observe that label assignment is stable for static actions, independent of the number of people being in front of the camera. One could assume that the stability changes in the dynamic condition depending on the number of people. Looking at the two corresponding videos in more detail, however, we identified that the type of dynamic actions differs between the two recordings. In this first case (cf. Figure 2a), dynamic actions did not lead to overlapping bounding boxes, while this was the case for the second trial. Thus, the descriptors that the underlying architecture extracts may contain information of multiple identities.



(a) 2, Close, Flat, Dynamic (b) 3, Close, Flat, Static

**Fig. 2**: Extracts from video footage corresponding to the bold trials in Table 3 for subjective analysis.

## 7. CONCLUSION

A universal method was presented for error resilient multi-shot gallery building in the open-set person re-identification. The gallery stores descriptors produced by an underlying architecture that allows measuring descriptor difference using L2 Euclidean distance. Unknown individuals are identified through outlier detection on the set containing the descriptors of a potential match from the gallery along with the query descriptor by using the Isolation Forest algorithm. Thus, there is no need to determine a threshold as in threshold based decision rules. For the quantitative analysis, we modified the evaluation set of the Market_1501 dataset to suit the open-set scenario. It was shown that our method significantly outperforms state-of-the-art approaches in terms of the TTR/FTR metrics. In terms of qualitative analysis, it was observed that the approach works sufficiently well in practice for controlled environments.

Finally, performance of the proposed method in extreme cases, e.g. long run time, larger number of identities and different environments can be investigated. Additionally, performance with other underlying architectures can be assessed.

## 8. REFERENCES


[1] Mang Ye, Jianbing Shen, Gaojie Lin, Tao Xiang, Ling Shao, and Steven C.H. Hoi, "Deep learning for person re-identification: A survey and outlook," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2021.

[2] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang, "Transfer re-identification: From person to set-based verification," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 2650–2657.

[3] Svebor Karaman and Andrew D Bagdanov, "Identity inference: generalizing person re-identification scenarios," in *European Conference on Computer Vision*. Springer, 2012, pp. 443–452.

[4] Michela Farenzena, Loris Bazzani, Alessandro Perina, Vittorio Murino, and Marco Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2010, pp. 2360–2367.

[5] Loris Bazzani, Marco Cristani, Alessandro Perina, Michela Farenzena, and Vittorio Murino, "Multiple-shot person re-identification by hpe signature," in *20th International Conference on Pattern Recognition*. IEEE, 2010, pp. 1413–1416.

[6] Tong Xiao, Hongsheng Li, Wanli Ouyang, and Xiaogang Wang, "Learning deep feature representations with domain guided dropout for person re-identification," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1249–1258.

[7] Chih-Ting Liu, Chih-Wei Wu, Yu-Chiang Frank Wang, and Shao-Yi Chien, "Spatially and temporally efficient non-local attention network for video-based person re-identification," *arXiv preprint arXiv:1908.01683*, 2019.

[8] Niall McLaughlin, Jesus Martinez Del Rincon, and Paul Miller, "Recurrent convolutional network for video-based person re-identification," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1325–1334.

[9] Alexander Hermans, Lucas Beyer, and Bastian Leibe, "In defense of the triplet loss for person re-identification," *arXiv preprint arXiv:1703.07737*, 2017.

[10] Shengyong Ding, Liang Lin, Guangrun Wang, and Hongyang Chao, "Deep feature learning with relative distance comparison for person re-identification," *Pattern Recognition*, vol. 48, no. 10, pp. 2993–3003, 2015.

[11] Wei-Shi Zheng, Xiang Li, Tao Xiang, Shengcai Liao, Jianhuang Lai, and Shaogang Gong, "Partial person re-identification," in *IEEE International Conference on Computer Vision*, 2015, pp. 4678–4686.

[12] Qingming Leng, Mang Ye, and Qi Tian, "A survey of open-world person re-identification," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 4, pp. 1092–1108, 2019.

[13] Madhawa Vidanapathirana, Imesha Sudasingha, Pasindu Kanchana, Jayan Vidanapathirana, and Indika Perera, "Open set person re-identification framework on closed set re-id systems," in *IEEE 2nd International Conference on Signal and Image Processing (ICSIP)*. IEEE, 2017, pp. 66–71.

[14] Xuan Zhang, Hao Luo, Xing Fan, Weilai Xiang, Yixiao Sun, Qiqi Xiao, Wei Jiang, Chi Zhang, and Jian Sun, "Alignedreid: Surpassing human-level performance in person re-identification," *arXiv preprint arXiv:1711.08184*, 2017.

[15] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou, "Isolation forest," in *Eighth IEEE International Conference on Data Mining*. IEEE, 2008, pp. 413–422.

[16] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang, "Towards open-world person re-identification by one-shot group-based verification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 3, pp. 591–606, 2016.

[17] Xiang Li, Ancong Wu, and Wei-Shi Zheng, "Adversarial open-world person re-identification," in *European Conference on Computer Vision*, 2018, pp. 280–296.

[18] Zhedong Zheng, Liang Zheng, and Yi Yang, "Unlabeled samples generated by gan improve the person re-identification baseline in vitro," in *IEEE International Conference on Computer Vision*, 2017, pp. 3754–3762.

[19] Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, and Pascal Frossard, "Deepfool: a simple and accurate method to fool deep neural networks," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2574–2582.

[20] Joseph Redmon and Ali Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.