

Boosting Re-identification in the Ultra-running Scenario

Miguel Ángel Medina^a, Javier Lorenzo-Navarro^b, David Freire-Obregón^c,
Oliverio J. Santana^d, Daniel Hernández-Sosa^e and Modesto Castrillón Santana^f

SIANI, Universidad de Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain

Keywords: Computer Vision in Sports, Re-identification, Biometrics.

Abstract: In the context of ultra-running (longer than a marathon distance), whole-body based re-identification (ReId) state of the art approaches have reported moderated success due to the challenging unrestricted characteristics of the long-term scenario, as very different illuminations, accessories (backpacks, caps, sunglasses), and/or changes of clothes are present. In this paper, we explore the integration of two elements in the ReId process: 1) an additional biometric cue such as the face, and 2) the particular spatio-temporal context information present in these competitions. Preliminary results confirm the limited relevance of the facial cue in the (not high resolution) ReId scenario and the great benefits of the contextual information to reduce the gallery size and consequently improve the overall ReId performance.

1 INTRODUCTION

Since the early 1990s, with the appearance of ChampionChip¹, massive running event organizers have tools to control the presence of runners in specific locations along the course track. Timing systems provide an essential mechanism to automatically analyze each competition collected time stamps in order to verify whether a participant's performance is suspicious of course cutting. Those systems are nowadays an ever-present element of running competitions. However, timing systems do not control the *real* presence of runners but of the tag they are carrying. They neither verify whether the person remains the same throughout the course, nor if the runner identity matches the one of the registered person. Such circumstances pose unpleasant situations to organizers, mainly related to insurance policies and incorrect classification results, producing a bad user experience.

Runners identification is starting to attract the attention of the computer vision community. So far, the literature has mainly focused on the Racing Bib

Number (RBN) recognition problem (Ben-Ami et al., 2012). Indeed, this solution suffers from the identical drawback of tag-based systems because it does not tackle the person identification problem. In this sense, biometrics can aid in the runner identification problem. However, as far as we know, biometric cues such as facial or body appearance and gait, have rarely been applied in the participant recognition process (Penate-Sanchez et al., 2020; Wrońska et al., 2017; Choi et al., 2021).

This paper explores the integration of different cues to evaluate their feasibility to re-identify individuals in a real ultra-running competition. Early work considering the whole-body appearance (Penate-Sanchez et al., 2020) has reported poorer ReId results compared to the state-of-the-art (SOTA) in more extensive ReId benchmarks. In this regard, we evaluate the integration of facial appearance, whenever the face is detected, and the temporal coherence present in ultra-trail competitions. The latter aims to reduce the gallery size, and therefore increase the final ReId performance. For this purpose, the dataset and proposal used in (Penate-Sanchez et al., 2020) are adopted as the benchmark and baseline, respectively.

^a <https://orcid.org/0000-0001-6734-2257>

^b <https://orcid.org/0000-0002-2834-2067>

^c <https://orcid.org/0000-0003-2378-4277>

^d <https://orcid.org/0000-0001-7511-5783>

^e <https://orcid.org/0000-0003-3022-7698>

^f <https://orcid.org/0000-0002-8673-2725>

¹ <https://www.mylaps.com/about-us/#our-story>

2 IDENTIFICATION IN RUNNING EVENTS

Computer vision in sports is an active field as suggested by recent surveys (Moeslund et al., 2014; Thomas et al., 2017) and workshops such as the CVPR's CVsports and ACM's MMSports. In those venues, the community has mainly focused on analyzing athlete movements and the study of team sports to collect statistical data. In both cases, to evidence ways of improvement (Li and Zhang, 2019; Thomas et al., 2017). As we are interested in participants ReId in a running event, our focus is diverse. We briefly summarise those proposals of which we are aware:

RBN Recognition. This is the primary adopted approach for runners identification or ReId. Even when these competitions are indeed organized nowadays almost everywhere, the number of publicly available data is limited. The literature on RBN recognition is not vast and mainly focuses on the Marathon context, characterized by daylight conditions and big fonts. The pioneering work by Ben Ami et al. (Ben-Ami et al., 2012) encloses a public dataset for RBN detection and recognition. Firstly face detection is applied to estimate the RBN Region of Interest (RoI), to later perform Optical Character Recognition (OCR). The initial face detection step is also adopted in (Boonsim, 2018; Shivakumara et al., 2017) to then apply text detection and recognition. The work described in (de Jesús and Borges, 2018) skips the face or body detection step, which may be fooled by the runner's poses, focusing on text detection.

More recently, deep learning has been adopted for RBN recognition (Kamlesh et al., 2017; Minghui Liao and Bai, 2018; Nag et al., 2019; Wong et al., 2019). In most proposals, an initial convolutional neural network (CNN) is used to detect the RBN, including a convolutional recurrent neural network (CRNN) for RBN recognition in a second stage.

Biometrics. Although body information has been used to determine the likely RBN location; the face, body, or clothing appearance has rarely been considered to identify participants in running competitions. In the work described in (Wrońska et al., 2017), the authors combine facial appearance with RBN recognition, improving the overall performance. This multimodal strategy is well suited, as face and RBN occlusions are frequently present, and the target pose may introduce difficulties to both face or RBN detection. Indeed, with daylight conditions, the marathon-like scenario reduces the wildness presence in the problem dataset. Bearing this in mind,

long-term ReId in the ultra-running scenario proposed in (Penate-Sanchez et al., 2020) defines a new benchmark for SOTA ReId approaches based on the body/clothing appearance. The final evaluation of top-ranked ReId approaches suggests the scenario difficulties. For this reason, we aim to assess the integration of facial information and temporal coherence in the loop.

Compared to the whole-body (WB), the face presents some advantages due to the trait permanence, particularly in scenarios covering a significant time lapse. For instance, participants may change their clothes along the track in ultra-running competitions. However, face ReId in low-resolution imagery shows significantly worse results compared to face recognition (FR) techniques in other scenarios. As pointed out in (Cheng et al., 2020), there are differences present in the classical surveillance scenario compared to the common FR context, further evidenced by the reduced success obtained by present face-based techniques in challenging benchmarks. This is also shown in the recent work (Dietlmeier et al., 2020), which suggests the reduced negative effect of blurring faces in ReId benchmarks in terms of overall system accuracy.

Even if those are not encouraging conclusions, we explore face integration in the selected scenario. For that purpose, a reliable face detector must be adopted. In recent years, some face detectors have been widely used by the community. The *dlib* (King, 2009) implementation of the Kazemi and Sullivan face detector (Kazemi and Sullivan, 2014) is one of them. With the recent irruption of deep learning, the latest face detectors have reduced appearance restrictions. A frequently used detector is the Multi-Task Cascaded Convolutional Networks (MTCNN) (Zhang et al., 2016). However, the problem is not completely solved given current challenges in face detection: 1) intra-class variation, 2) face occlusion, and 3) multi-scale face detection. Among the publicly available top ranked detectors, we may mention: Tiny face detector (Hu and Ramanan, 2017), SSH (Najibi et al., 2017), FaceKit (Shi et al., 2018), S³FD (Zhang et al., 2017) and RetinaFace (Deng et al., 2019). RetinaFace in particular claims to provide SOTA results in FDDB (Jain and Learned-Miller., 2010) and WiderFace (Yang et al., 2016) datasets.

Previous proposals focus on photographs or extracted video frames but do not integrate temporal information. In this sense, we may mention the recent CampusRun dataset (Napolean et al., 2019) of images captured for each participant every kilometer by hand-held cameras during a half marathon event, offering a large number of samples per identity. Given the video

stream availability, the gait trait is adopted in the work described in (Choi et al., 2021) focusing on the arm swing features extracted from the silhouette. The authors strategy is to remove the problems present in RBN or face occlusion, or the similarity in clothing appearance (same team).

3 PROPOSED METHOD

3.1 Whole-body and Face ReId

In the ReId dataset proposed in (Penate-Sanchez et al., 2020), runners' WB bounding boxes are provided. That work evaluated two recent WB person ReId approaches: AlignedReID++ (Luo et al., 2019) and ABD (Chen et al., 2019), after their SOTA results in Market-1501 (Zheng et al., 2015). In light of the conclusions of (Penate-Sanchez et al., 2020), WB descriptors will be computed with AlignedReID++, since it exhibited slightly better results.

For face processing, we have adopted a suitable face detector for low-resolution images (RetinaFace) since it performs better than other face detectors in the comparison reported below (see Section 4.2). Once the face is detected, since faces may be frequently captured at too low resolution for robust FR, integrating a superresolution preprocess (SR) is evaluated. In this regard, the cropped faces are represented at higher resolution, making use of the Local Implicit Image Function (LIIF) (Chen et al., 2021), exploring the use of different representation resolutions to define the best one suited for the scenario.

Later, facial embeddings are obtained with VG-GFace2 (Cao et al., 2018) that suggests being the best choice according to our results, after making an extensive analysis using the Deepface framework (Serengil, 2021), which wraps a collection of SOTA FR models.

When both cues, WB and face, are combined, in the reported first experiments, a feature level fusion is carried out with the WB mentioned above and face descriptors, e.g., AlignedReID++ and VGGFace2, respectively.

3.2 Temporal Coherence

An ultra-running competition has a clear timeline, as participants reach each aid station in order, not repeating any of them. Assuming that the proposed vision-based system would be working in parallel with traditional timing systems, runners elapsed times at each specific location are available. Moreover, runners are also captured at these locations. For this reason, these

specific locations across the paper are known below as Recording Points (RPs).

The classical experimental evaluation in ReId aims to determine which person in a gallery matches the probe's identity. In this regard, we could consider an individual detected in RP_{k+1} as probe, and all the samples from RP_k as the gallery. Thus, for a particular RP_k , where $k = \{1, \dots, r\}$ in the evaluated dataset, there are a number of participants recorded, n_k , corresponding $p_{i,k}$ to the participant crossing in position i th through RP_k , $b_{i,k}$ his/her bib number, and $t_{i,k}$ his/her elapsed time at RP_k . Given the probe sample $p_{i,k+1}$ in RP_{k+1} , its gallery in RP_k is defined as:

$$\text{gallery} = \{p_{j,k}\} \text{ for } j = 1, \dots, n_k \quad (1)$$

Including the whole set of identities captured in RP_k as gallery is not smart. As illustrated in Figure 1 when first runners arrive to RP_{k+1} some runners have still not crossed through RP_k . Therefore, a simple gallery temporal coherence filtering would be to remove for a given probe, captured at a particular timestamp, those gallery samples that have not yet reached RP_k . But indeed, given the spatial distances between RPs, any runner would need some time to reach RP_{k+1} physically. Thus, we can apply more restrictive filtering with something like *runners should have passed RP_k at least x minutes ago*. To define x , a possible approach could be to use the organization's expected time between RPs, which is estimated in advance based on the distance and accumulated positive slope between both RPs. However, weather conditions may also affect runners performance. For that reason, we have adopted a strategy that uses the actual runners performance on the particular competition date. In this sense, we will take into account the time difference between the first runner to cross both RPs, $p_{1,k}$ and $p_{1,k+1}$:

$$\Delta t_{1,k+1} = t_{1,k+1} - t_{1,k} \quad (2)$$

This value would be the x mentioned above, and it could be used to filter the gallery defined in Equation (1), assuming that any runner should need at least that time to reach RP_{k+1} .

$$\text{gallery} = \{p_{j,k}\} \text{ such that } t_{j,k} < t_{i,k+1} - (\Delta t_{1,k+1}) \quad (3)$$

However, that value may be risky, as runners could exchange positions. We illustrate this circumstance with an example. Consider that the first three runners depicted in Figure 1 (red, blue and green t-shirts) passed RP_k at time 00:05:00, 00:05:20 and 00:06:30, respectively. The same three runners led the race in RP_{k+1} but exchanged positions (green, blue and red), passing RP_{k+1} at time

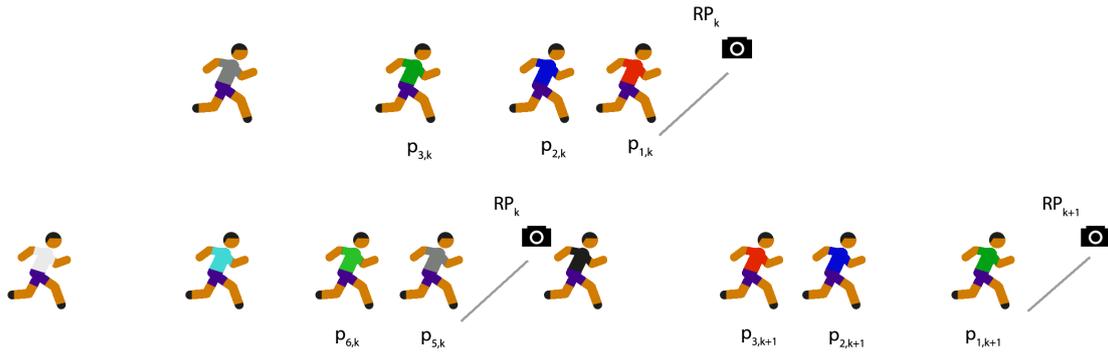


Figure 1: Temporal coherence illustration. Top) Let’s assume that red, blue, and green runners cross RP_k in that order. Bottom) Red and green runners exchange positions in RP_{k+1} when they arrive roughly one hour later. Using the temporal coherence, for a runner arriving at RP_{k+1} , we remove from the gallery those individuals who have not had enough time to reach RP_{k+1} from RP_k . The runner position estimates the time threshold and attenuates to consider the possibility of runners performance drops.

01:15:32, 01:15:50 and 01:16:30, respectively. In such case, $\Delta t_{1,k+1} = 01 : 10 : 32$. If we are considering the gallery set for the first runner probe, $p_{1,k+1}$ then according to the rule of Equation (3), candidate runners in the gallery should verify $t_{j,k} < t_{1,k+1} - (\Delta t_{1,k+1})$ which fixes the temporal threshold in RP_k to $01 : 15 : 32 - 01 : 10 : 32 = 00 : 05 : 00$ excluding runners blue and green from the gallery, i.e. the desired runner will be excluded from his own gallery. To overcome that circumstance, forced by the runners performance variations along the track, the temporal threshold is slightly weighted, introducing thr , to avoid removing an excessive number of identities from the gallery.

$$gallery = \{p_{j,k}\} \text{ such that } t_{j,k} < t_{i,k+1} - (\Delta t_{1,k+1} \times thr) \quad (4)$$

where $thr \in (0, 1]$. Defining its value to 1.0, it would be equivalent to Equation (1) where it is likely that some runners may be excluded from their gallery. Assigning a lower value will define a margin to skip those situations.

The previous idea explains the overall gallery filtering procedure. However, we have also explored using a variable threshold due to the conservative nature of a fixed threshold defined by the top runners’ performance. Thus, we introduce additional restrictions in the gallery building. Given that the fastest split time is not reachable for most runners, we define an adaptive split time threshold that considers the different runners’ performances. For any two previous consecutive RPs, RP_{l-1} and RP_l , where $(l < k)$, for a particular runner in position i , the RPs split difference may be computed similarly as:

$$\Delta t_{i,l} = t_{i,l} - t_{i,l-1} \quad (5)$$

Given elapsed times in previous RP_s ($1, \dots, k -$

1), they serve to estimate the split time for runner in position i between RP_{k-1} and RP_k as:

$$\widetilde{\Delta t_{i,k}} = w_0 + w_1 \Delta t_{i,2} + \dots + w_{k-2} \Delta t_{i,k-1} \quad (6)$$

Our proposal evaluated such estimation using a linear regression (LR) model and a Random Forest (RF) model. The latter reported better results in the experiments presented below and therefore was adopted to compute the adaptive threshold to filter the gallery for a particular $p_{i,k+1}$. In summary, the adaptive filtering is expressed as:

$$gallery = \{p_{j,k}\} \text{ such that } t_{j,k} < t_{i,k+1} - (\widetilde{\Delta t_{i,k+1}} \times thr) \quad (7)$$

In the experiments below, the margin thr introduced to cope with the error in the estimated split difference using Equation (6) is defined to 0.9, avoiding the artifact of removing from the gallery the runner that corresponds to the probe sample $p_{i,k+1}$. This value worked adequately in the dataset where RPs are located at least ten kilometers away from each other, with the fastest runners requiring almost one hour to cover the distance.

Using Equation (7) to determine the gallery for a probe, will have the effect of increasing the gallery size according to runner position. Indeed, the gallery corresponding to the last runners would comprise all identities that passed RP_k , as any runner passed that point enough time ago. Therefore, given that the existing timing systems provide information about runners that have already passed a particular RP, we imposed a second rule in the gallery filtering procedure. This second rule removes from the gallery those samples belonging to bib numbers who have already left RP_{k+1} . Therefore, for participant $p_{i,k+1}$ his/her gallery is defined as:

$$gallery = \{p_{j,k}\} \text{ that } \begin{cases} t_{j,k} < t_{i,k+1} - (\widetilde{\Delta t_{i,k+1}} \times thr) \\ b_{j,k} \notin b_{r,k+1} \text{ for } r = 1, \dots, i-1 \end{cases} \quad (8)$$

To summarize, we use the split time (elapsed time difference between two consecutive RPs) to filter the samples present in the ReId gallery. Two rules are applied to remove samples corresponding to bib numbers: 1) who have physically not been able to reach the current RP, and 2) who have already left the current RP. The temporal coherence strategy combining both restrictions, adaptive threshold, and already matched bibs removal are referred to in the experiments below as TC.



Figure 2: Transgrancanaria 2020 track with RPs marked as red circles (source image from Google Maps).



Figure 3: An RP setup.

4 EXPERIMENTAL EVALUATION

4.1 Dataset Description

As mentioned above, this paper evaluates the combination of different cues to improve runner's ReId in footage captured during a running event. We make use of the annotated dataset described in (Penate-Sanchez et al., 2020), a challenging ReId benchmark with just 109 identities. The dataset was collected during Transgrancanaria (TGC) Classic 128KM 2020, where runners departed on March 6, 2020, at 11 pm, closing the finish line 30 hours later. TGC participants were recorded at different locations along the race track, see Figure 2, using a set of Sony Alpha ILCE 6400 (16-50mm lens) configured at 50 fps and 1920×1080 pixels resolution, see Figure 3. TGC participants were later annotated in five different RPs (RP1-RP5). The capture conditions vary significantly among the different RPs, as evidenced in Figure 4.

The dataset annotation comprises body containers at 1fps and RBN information extracted from the competition's official classification results. Table 1 summarizes the dataset statistics. The reader may observe that the closer to the finish line is the RP, the lower the number of annotated participants, even if the number of captured frames is larger. Indeed, not every runner reaches the finish line, and the elapsed time between leaders and the group tail increases significantly, e.g., leaders pass through RP5 approximately 13 hours before the last runners.

4.2 Face Detection Results

Even if different recent works (Cheng et al., 2020; Dietlmeier et al., 2020) have reported the low relevance of facial information in the surveillance/ReId scenario, we evaluate FR feasibility in the dataset.

A necessary previous step is face detection. We have evaluated three detectors: 1) The dlib (King, 2009) implementation of the Kazemi and Sullivan detector (Kazemi and Sullivan, 2014) denoted below as DLIBHOG, 2) MTCNN (Zhang et al., 2016), and 3) RetinaFace (Deng et al., 2019). Face detection results in Table 2 include the respective numbers of images processed, participant's bounding boxes, and recall.

The reported results clearly show the limited detector's performance in the scenario, evidencing its difficulties, particularly under nightlight conditions. DLIBHOG cannot manage the challenging low resolutions and poses present in the dataset. MTCNN behaves better, with an evident problem in nightlight images. RetinaFace can provide valid positive detections



Figure 4: Leaders of the TGC 2020 Classic recorded at the different RPs. Images from (Penate-Sanchez et al., 2020).

Table 1: TGC20ReId dataset statistics (Penate-Sanchez et al., 2020). RP1-2 are captured with nightlight.

Location	Km	Start Rec. Time	Original footage (frames)	# ids annot.
RP1	16.5	00:06	140,616	419
RP2	27.9	01:08	432,624	586
RP3	84.2	07:50	667,872	203
RP4	110.5	10:20	1,001,208	139
RP5	124.5	11:20	1,462,056	114

Table 2: Face detection results in terms of true positive rate (TPR) for a number of annotated runners (BBs). The number of true positives (TP) is also provided.

RP	# imgs	# BBs	DLIBHOG	MTCNN	RetinaFace
			TPR	TPR	TPR
RP1	1172	1589	0.00	0.03	0.19
RP2	1234	1445	0.02	0.07	0.45
RP3	618	526	0.01	0.35	0.61
RP4	281	255	0.02	0.39	0.84
RP5	250	253	0.03	0.45	0.75
Total	3555	4068	0.01	0.13	0.41

Table 3: ReId results using samples captured in RP5 as probe and the gallery the corresponding captures in RP4 not integrating TC.

Cue	CMC		mAP
	rank-1	rank-5	
WB (Penate-Sanchez et al., 2020)	43.15	75.51	48.37
Face	5.14	13.97	10.21
Face+SR	7.35	14.70	11.13
WB+Face	45.22	75.51	49.19

rate roughly 60-80% in daylight RPs, and even over 45% in RP2, while hardly 20% in RP1. The latter’s performance might be justified due to the frequent presence of headlamps in RP1, making face visibility

quite challenging. In any case, RetinaFace seems to be good enough, with an overall detection rate of over 40% (note that almost 75% of the annotated participants were captured with nightlight). Daylight detec-

Table 4: ReId results using samples captured in RP5 as probe and the gallery the corresponding captures in RP4 integrating TC.

Cue	CMC		mAP
	rank-1	rank-5	
WB+TC	78.84	91.29	78.69
Face+SR	7.35	14.70	11.13
Face+SR+TC	18.68	50.50	29.11
WB+Face+TC	78.84	91.29	78.52
WB+Face+SR+TC	78.00	91.29	77.82

tions are the *easiest* ones for facial detectors, as being the most similar ones present in the training sets of the detection methods.

Using RetinaFace, the mean and standard deviation of face containers in RP5 is $(127.5 \pm 76.2 \times 102.7 \pm 85.2)$ with a minimum detected face of (8×10) pixels. Therefore, some extracted faces are certainly low resolutions ones.

4.3 ReId Results

As mentioned above, the classical experimental scenario in ReId aims to determine which person in a gallery matches the probe’s identity. In this regard, we could consider an individual detected in R_{k+1} as probe, and all the samples from R_k as the gallery. For the experiments, we will use standard metrics which are well established in recent Re-ID papers (Luo et al., 2019; Chen et al., 2019).

The cumulative matching characteristics curve (CMC) ranks the gallery samples according to the distance to the probe. For the given probe, the CMC top-k accuracy is 100% if the first k ranked gallery samples contain the probe identity, and 0 otherwise. The final CMC curve averages the respective ‘probe curves. In summary, a CMC rank-1 of 80% indicates that the correct identity is ranked first for 80% of the probes. When multiple instances of the identity are present in both query and gallery sets, the mean average precision scores (mAP) are better suited. For a p number of probes, mAP is defined as $mAP = \frac{\sum_{i=1}^p AP_i}{p}$, where AP_i refers to the area under the precision-recall curve of probe i .

The purpose is to make use of standard ReId metrics, being able to compare to the closed-set ReId evaluation protocol presented for the same dataset in (Penate-Sanchez et al., 2020). We firstly analyze separately individual cues and their possible combinations, focusing on the more favorable situations. Consequently, the last RPs where daylight conditions ease face detections and runners elapsed time allow the system to take further advantage of TC to reduce the gallery size.

Table 3 summarizes results achieved considering RP5 samples as probe and RP4 as gallery. Both cosine and euclidean distances have been evaluated, with similar results, presenting just the former. The Deep-face framework has been used for faces, including the best results that correspond to VGGFace2. Specifically for faces, different SR resolutions have been explored, presenting results for 100×100 facial bounding boxes. Observing the table in detail, on the one side, it is evident that the use of the facial pattern does not present a performance similar to the one exhibited by the WB. Indeed, the results are pretty poor even if SR preprocessing is applied. Those results suggest that faces are not yet a valid cue for our purpose in this scenario.

The results after integrating SR and/or TC are summarized in Table 4. The comparison with Table 3 suggests that applying TC improves the performance, as it reduces the gallery size. For instance, considering RP5 samples as the probe and RP4 as the gallery, the closed-set ReId benchmark for 109 identities contains 208 samples without TC. With TC, the resulting RP4 average gallery size is 28. TC significantly improves both WB and facial ReId results due to this gallery reduction. Nevertheless, the facial results are considerably worse than those achieved using just WB features. Finally, WB and face fusion do not report an overall improvement.

In summary, the results reported in Table 4 evidence that the combination WB+TC leads the charts. Those results are achieved for two specific RPs, mostly under daylight capture conditions and closer to the finish line, i.e. with a large temporal difference among participants. To provide a different view of the TC influence, Table 5 summarizes for any two RPs the results obtained with and without TC. Those results verify the TC approach robustness, with the expected more substantial influence for last RPs where the elapsed time differences among runners are significantly larger.

5 CONCLUSIONS

This paper has explored existing ReId proposals in the ultra-running scenario. In addition to WB appearance, which is insufficient for medium or long-term ReId, we have evaluated the facial trait, with higher permanence, and the integration of TC in a context where clothes variations are possible.

The reported results suggest the following: 1) the poor performance of applying FR techniques for this surveillance/ReId scenario, and 2) the benefits of the TC strategy that takes advantage of the presence of

Table 5: Top) Rank-5 and bottom) mAP results were computed densely between all RPs (WB/WB+TC). When the gallery belongs to a previous RP, the second rule is not applied, and the sense of the eq. 5 is modified, and the estimated split time is added instead of subtracting.

		Probe				
		RP1	RP2	RP3	RP4	RP5
Gallery	RP1	-	32.4/36.2	20.2/32.8	46.1/57.8	47.3/65.2
	RP2	32.7/33.5	-	11.8/31.9	27.9/51.5	37.3/52.7
	RP3	28.3/29.3	15.7/32.4	-	33.3/53.4	25.3/44.0
	RP4	42.9/42.9	23.2/45.4	26.9/42.9	-	75.5/91.3
	RP5	23.3/23.8	18.4/30.8	7.98/42.9	55.8/85.8	-

		Probe				
		RP1	RP2	RP3	RP4	RP5
Gallery	RP1	-	15.0/17.9	9.2/16.0	23.0/36.3	22.7/36.8
	RP21	20.4/21.8	-	7.22/10.3	13.6/28.2	20.6/33.0
	RP3	12.3/12.4	10.0/21.4	-	19.5/39.2	15.2/28.4
	RP4	28.4/28.5	11.3/25.2	16.6/29.3	-	48.4/78.7
	RP5	16.2/16.7	11.1/21.8	7.3/28.9	38.0/69.6	-

a tag-based time control system. The former has already been pointed out in SOTA by different authors (Cheng et al., 2020; Dietlmeier et al., 2020). These authors stated that the benefits introduced by the integration of the facial cue are reduced, even if SR is adopted to increase the face sample resolution. The latter opportunistically uses additional information present in the process, significantly improving ReId metrics for the challenging dataset.

In any case, the challenging scenario posed by the dataset is evident, where rank-1 CMC does not even reach 80%. There is plenty of room for improvement, as there are multiple aspects to explore. The results suggest the benefit of using WB, but in the long-term, a scenario where clothes may change makes it unfeasible to use WB appearance as a single cue for ReId. It will undoubtedly be fooled by clothes changes along the track. Therefore, additional cues are necessary. Our results indicate that current FR techniques do not improve the overall performance, but other strategies, such as TC, may help.

ACKNOWLEDGEMENTS

This work is partially funded by the ULPGC under project ULPGC2018-08, the Spanish Ministry of Economy and Competitiveness (MINECO) under project RTI2018-093337-B-I00, the Spanish Ministry of Science and Innovation under project PID2019-107228RB-I00, and by the Gobierno de Canarias and FEDER funds under project ProID2020010024.

REFERENCES

- Ben-Ami, I., (Basha), T. D., and Avidan, S. (2012). Racing bib number recognition. In *British Machine Vision Conference*, pages 1–10, Surrey, UK. British Machine Vision Association.
- Boonsim, N. (2018). Racing bib number localization on complex backgrounds. *WSEAS Transactions on Systems and Control*, 13:226–231.
- Cao, Q., Shen, L., Xie, W., Parkhi, O. M., and Zisserman, A. (2018). VGGFace2: A dataset for recognising faces across pose and age. In *International Conference on Automatic Face and Gesture Recognition*.
- Chen, T., Ding, S., and Wang, Z. (2019). ABD-Net: Attentive but diverse person re-identification. In *The IEEE International Conference on Computer Vision (ICCV)*.
- Chen, Y., Liu, S., and Wang, X. (2021). Learning continuous image representation with local implicit image function. In *Proceedings of the International Conference on Computer Vision*.
- Cheng, Z., Zhu, X., and Gong, S. (2020). Face re-identification challenge: Are face recognition models good enough? *Pattern Recognition*, 107:107422.
- Choi, Y., Napoleon, Y., and van Gemert, J. C. (2021). The arm-swing is discriminative in video gait recognition for athlete re-identification.
- de Jesús, W. M. and Borges, D. L. (2018). An improved stroke width transform to detect race bib numbers. In *Proceedings of the Mexican Conference on Pattern Recognition*, pages 267–276, Puebla, Mexico. Springer.
- Deng, J., Guo, J., Zhou, Y., Yu, J., Kotsia, I., and Zafeiriou, S. (2019). Retinaface: Single-stage dense face localisation in the wild. *arXiv*, 1905.00641.
- Dietlmeier, J., Antony, J., Mcguinness, K., and O’Connor, N. E. (2020). How important are faces for person re-identification? In *Proceedings International Conference on Pattern Recognition*, Milan, Italy.

- Hu, P. and Ramanan, D. (2017). Finding tiny faces. In *Proceedings of Computer Vision and Pattern Recognition*, pages 951–959, Hawaii, USA. IEEE.
- Jain, V. and Learned-Miller, E. (2010). Fddb: A benchmark for face detection in unconstrained settings. Technical report, University of Massachusetts, Amherst.
- Kamlesh, P. X., Yang, Y., and Xu, Y. (2017). Person re-identification with end-to-end scene text recognition. In *Chinese Conference on Computer Vision*, pages 363–374, Tianjin, China. Springer.
- Kazemi, V. and Sullivan, J. (2014). One millisecond face alignment with an ensemble of regression trees. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1867–1874, Columbus, OH, USA. IEEE.
- King, D. E. (2009). Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10:1755–1758.
- Li, G. and Zhang, C. (2019). Automatic detection technology of sports athletes based on image recognition technology. *EURASIP Journal on Image and Video Processing*, 2019(1):15.
- Luo, H., Jiang, W., Zhang, X., Fan, X., Qian, J., and Zhang, C. (2019). AlignedReID++: Dynamically matching local information for person re-identification. *Pattern Recognition*, 94:53–61.
- Minghui Liao, B. S. and Bai, X. (2018). TextBoxes++: A single-shot oriented scene text detector. *IEEE Transactions on Image Processing*, 27(8):3676–3690.
- Moeslund, T. B., Thomas, G., and Hilton, A. (2014). *Computer Vision in Sports*. Springer, Switzerland.
- Nag, S., Ramachandra, R., Shivakumara, P., Pal, U., Lu, T., and Kankanhall, M. (2019). Crnn based jersey-bib number/text recognition in sports and marathon images. In *International Conference on Document Analysis and Recognition (ICDAR)*, pages 1149–1156, Sydney, Australia. IEEE.
- Najibi, M., Samangouei, P., Chellappa, R., and Davis, L. S. (2017). SSH: Single stage headless face detector. In *Proceedings of IEEE International Conference on Computer Vision*, pages 4875–4884, Venice, Italy. IEEE.
- Napolean, Y., Wibow, P. T., and van Gemert, J. C. (2019). Running event visualization using videos from multiple cameras. In *ACM International Workshop on Multimedia Content Analysis in Sports*.
- Penate-Sanchez, A., Freire-Obregón, D., Lorenzo-Melián, A., Lorenzo-Navarro, J., and Castrillón-Santana, M. (2020). TGC20ReId: A dataset for sport event re-identification in the wild. *Pattern Recognition Letters*, 138:355–361.
- Serengil, S. I. (2021). deepface: A lightweight face recognition and facial attribute analysis framework for python. <https://github.com/serengil/deepface>.
- Shi, X., Shan, S., Kan, M., Wu, S., and Chen, X. (2018). Real-time rotation-invariant face detection with progressive calibration networks. *arXiv*, 1804.06039.
- Shivakumara, P., Raghavendra, R., Qin, L., B.Raja, K., Luc, T., and Pal, U. (2017). A new multi-modal approach to bib number/text detection and recognition in marathon images. *Pattern Recognition*, 61:479–491.
- Thomas, G., Gade, R., Moeslund, T. B., Carr, P., and Hilton, A. (2017). Computer vision for sports: Current applications and research topics. *Computer Vision and Image Understanding*, 159:3–18.
- Wong, Y. C., Choi, L. J., Singh, R. S. S., Zhang, H., and Syafeeza, A. R. (2019). Deep learning based racing bib number detection and recognition. *Jordanian Journal of Computers and Information Technology (JJCIT)*, 5(3):181–194.
- Wrońska, A., Sarnacki, K., and Saeed, K. (2017). Athlete number detection on the basis of their face images. In *Proceedings International Conference on Biometrics and Kansei Engineering*, pages 84–89, Kyoto, Japan. IEEE.
- Yang, S., Luo, P., Loy, C.-C., and Tang, X. (2016). Wider face: A face detection benchmark. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5525–5533, Hawaii, USA. IEEE.
- Zhang, K., Zhang, Z., Li, Z., and Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503.
- Zhang, S., Zhu, X., Lei, Z., Shi, H., Wang, X., and Li, S. Z. (2017). S³FD: single shot scale-invariant face detector. In *IEEE International Conference on Computer Vision (ICCV)*, pages 192–201, Venice, Italy. IEEE.
- Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., and Tian, Q. (2015). Scalable person re-identification: A benchmark. In *Proceedings of the International Conference on Computer Vision*.