View Reviews

Paper ID
950

Paper Title
PeR-ViS: Person Retrieval in Video Surveillance using Semantic Description

Track Name
Second Round Submission

Reviewer #1

Not Submitted

Reviewer #3

Questions

1. PAPER SUMMARY What is the paper about? Please, be concise (2 to 3 sentences).
The paper presents an automated pedestrian re-identification (retrieval) scheme in a video surveillance context using gender, physical and appearance attributes. The paper claims to use "semantic description" for retrieving specific pedestrian instances. Experiments are performed on a surveillance-task-oriented public dataset.

2. PAPER STRENGTHS Please discuss, justifying your comments with the appropriate level of details, the strengths of the paper (i.e. novelty, theoretical approach and/or technical correctness, adequate evaluation, clarity, etc).
The paper has unfortunately no strong aspects in my view.

3. PAPER WEAKNESSES Please discuss, justifying your comments with the appropriate level of details, the weaknesses of the paper (i.e. lack of novelty – given references to prior work-, lack of novelty, technical errors, or/and insufficient evaluation, etc). Note: If you think there is an error in the paper, please explain why it is an error.
The paper suffers from many weak issues.

Missing novelty: it shows a very large content overlap (concept, text, illustration) with reference 4. It seems to me that some parts of Ref. 4 were replaced by somewhat more modern modules (DenseNet instead of AlexNet)

Lacking experimental validation: as the topic strongly overlaps with pedestrian re-id, it would be imperative to assess it in terms of public datasets and common metric (e.g. CMC curve). As learned appearance representation plays an important role in the current work, it is difficult to judge its quality without having extensive testing,

Dataset should be also analyzed to give an impression of the complexity of a retrieval task.

Methodology contains much heuristics, e.g. torso/leg region estimation.

The paper's narrative is difficult to folow, in my view. I did not fully understand the claim "semantic description". Are named colors used? Some figures are not very informative (Figure 4 - Figure 7)- Figure 4-5. one needs to know the dataset well to be able to interpret this plots.

Mostly due to the limited novelty (overlap with [4]) and poor methodology [simple chain plugged together an applied to a simplistic dataset], I advide a definite reject. The paper could be much improved by broadening its scope and adopting a more analytic view on datasets, method components and experimental validation.

4. RECOMMENDATION
Strong Reject

5. JUSTIFICATION Justify your recommendation based on the strengths and weaknesses. Please be considerate to the authors and provide constructive feedback.
Mostly due to the limited novelty (overlap with [4]) and poor methodology [simple chain plugged together an applied to a simplistic dataset], I advide a definite reject.

Reviewer #4

Questions

1. PAPER SUMMARY What is the paper about? Please, be concise (2 to 3 sentences).
The paper proposes to use the semantic description of a subject to retrieve specific attributes such as height, gender, etc. The framework is composed of a Mask-RCNN network to perform semantic segmentation, followed by a DenseNet model to perform soft-biometric classification.

2. PAPER STRENGTHS Please discuss, justifying your comments with the appropriate level of details, the strengths of the paper (i.e. novelty, theoretical approach and/or technical correctness, adequate evaluation, clarity, etc).
The concept of using semantic description instead of visual data is intriguing. The results shown in Table 1 and 2 are quite interesting and infomative.
The experiments results are well documented and elaborated.

3. PAPER WEAKNESSES Please discuss, justifying your comments with the appropriate level of details, the weaknesses of the paper (i.e. lack of novelty – given references to prior work-, lack of novelty, technical errors, or/and insufficient evaluation, etc). Note: If you think there is an error in the paper, please explain why it is an error.

The novelty of the paper is limited. It consists of well known modules such as Mask R-CNN and DenseNet, which have already been used in previous papers. In fact, [1] has used exactly the same setup for the task. Additionally, [2] also employs semantic description, but uses AlexNet instead of DenseNet.

The pipeline followed by the paper is identical to that of [1], where the Mask-RCNN performs instance segmentation followed by DenseNet.

Certain parts of the paper seems to be overselling. For instance, in the introduction, I would not consider the Mask RCNN to be a major contribution of the paper, since it has already been proposed previously for the task of instance segmentation.

[1]Yaguchi, Takuya, and Mark S. Nixon. "Transfer learning based approach for semantic person retrieval." 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). IEEE, 2018.
[2]Galiyawala, Hiren, et al. "Person retrieval in surveillance video using height, color and gender." 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). IEEE, 2018.

4. RECOMMENDATION
Weak Reject

5. JUSTIFICATION Justify your recommendation based on the strengths and weaknesses. Please be considerate to the authors and provide constructive feedback.
Although the experimental results show merit, the paper is an incremental work over previous methods. In its current form, the novelty of the paper is limited and does not warrant an acceptance. I feel that employing semantic description was an interesting direction and could be pursued in more detail. Sticking to the already well explored pipeline of Mask R-CNN+DenseNet/AlexNet does not seem novel.

Reviewer #5

Questions

1. PAPER SUMMARY What is the paper about? Please, be concise (2 to 3 sentences).
The authors proposed a network for person retrieval in video that uses a filtering system for improved accuracy. They compare their method to several other methods and show that their method which uses person height, appearance, and gender for filtering achieves the best performance.

2. PAPER STRENGTHS Please discuss, justifying your comments with the appropriate level of details, the strengths of the paper (i.e. novelty, theoretical approach and/or technical correctness, adequate evaluation, clarity, etc).
+ Comparison to several other methods.
+ State-of-the-art performance on a common benchmark dataset.
+ Good qualitative results showing both success and failure cases in different scene complexity scenarios.
+ Breakdown showing effect of different descriptors on validation accuracy.

+ Clear description of the dataset and training parameters.
+ Ablation showing impact of different feature extraction backends.

3. PAPER WEAKNESSES Please discuss, justifying your comments with the appropriate level of details, the weaknesses of the paper (i.e. lack of novelty – given references to prior work-, lack of novelty, technical errors, or/and insufficient evaluation, etc). Note: If you think there is an error in the paper, please explain why it is an error.

- Several parts of the text are difficult to understand. For example, from lines 410-416, the authors talk about how some sequences have incorrect average height calculations. It is not clear what the "similar training video sequence" that has this error is, or why the proposed subtraction helps solve the problem.
- The proposed list of used descriptors seems arbitrary given the large amount of metadata provided in the dataset.
- It seems unfair to use the more sophisticated height feature for filtering while comparing to other methods that did not use such a strong feature. While the other methods could have used the height as computed in this work, they focus on other attributes that do not require such precise knowledge of the sensor geometry, making them potentially more transferable to other systems. Since filtering happens in discrete phases and begins with height, it is concerning that there is not evaluation for how IoU improves with each filtering stage. For example, if there is steady improvement with each stage then my concern would be alleviated, but I fear that most of the performance increase is coming from using more specific height information than is present in the dataset directly.

4. RECOMMENDATION

Borderline

5. JUSTIFICATION Justify your recommendation based on the strengths and weaknesses. Please be considerate to the authors and provide constructive feedback.

They achieve state-of-the-art results on the dataset used, but the metadata used for filtering is much different than what other methods use, making the comparisons to the more general methods difficult to interpret. It is unclear given the evaluation in the paper the method is better than others, or if height is simply a better filter. While the height was derived from information provided in the dataset, given that height estimates are not known in general, and the method used is this paper to generate them is not always applicable, I feel that this is a significant issue.