

View Reviews

Paper ID

430

Paper Title

PeR-VIS: Person Retrieval in Video Surveillance using Semantic Description

Track Name

First Round Submission

Reviewer #1

Questions

1. PAPER SUMMARY What is the paper about? Please, be concise (2 to 3 sentences).

The paper analyzes the problem of person retrieval in video surveillance with a semantic description. To solve this problem, authors propose a deep learning-based cascade filtering approach (PeR - ViS), which uses Mask R-CNN [6] (person detection and instance segmentation) and DenseNet - 161 [7] (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 paper is well written and easy to read. The proposed approach is interesting and well described. The state of the art is almost exhaustive.

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 main drawback of the paper is the limited novelty with respect to the state of the art. The paper is based on well-known techniques and a limited amount of new ideas is introduced. The only new contribution is the use of the person height. Moreover, the experimental protocol is very limited either in term of number of tests either in terms of comparisons with other approaches.

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.

The paper presents an interesting problem, but the proposed solution lacks of novelty. The paper does not introduce sufficient new ideas with respect to the state of the art. Moreover, the experimental protocol is very limited either in term of number of tests either in terms of comparisons with other approaches.

6. REVISION OPTION - QUESTION ONE If not accepted in this first round, should the paper be invited to resubmit a revised version of the paper? It would then be considered by the same reviewers and area chair.

REJECT Preferred. Authors will need to submit a SUBSTANTIALLY IMPROVED version for review.

7. REVISION OPTION - QUESTION TWO Please clearly indicate the issues that you feel should be addressed in a revised version such that the paper could potentially be considered for acceptance.

Authors should better point out the novelty of the paper. Moreover, the experimental protocol must be revised and enriched with additional tests.

Reviewer #3

Questions

1. PAPER SUMMARY What is the paper about? Please, be concise (2 to 3 sentences).

The paper is about person retrieval in video surveillance with some semantic description. A cascaded filter with some attribute conditions is used for get the retrieval results from the person candidates detected by Mask-R-CNN.

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).

pros:

1. The work attempt to execute person retrieval from original surveillance videos, which is a challenging problem to integrate several visual analysis modules, e.g., detection, height estimation, into an effective system.
2. Height is calculated with the help of camera calibration, which may be more accurate than direct height estimation from images.

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.

Cons:

1. The technical novelty in the paper is very limited. The key idea is a cascaded filter based retrieval method, however, how to decide the order of retrieval conditions in the cascade filter? It cannot be fixed for different query images, as someone maybe salient due to his red skirt, and someone is due to its bald head. And other techniques used in the papers, Mask-RCNN, attribute recognition, are not new. Little modification is proposed to improve for the specific use in surveillance scenes.
2. As the height is the first filter, the error or uncertainty in height estimation may have serious effects on the later retrieval process. For example, how to discrete the person height into retrieval conditions? 3 states? or 5 states? And how to handle the uncertainty in height description and height estimation? If the Average height is defined as the range 150cm to 170cm, a person with height estimation of 172cm will be decided as tall, but human user may still use Average condition to find him.
3. Compared to other related work on person search and attribute based retrieval, the papers just use a very limited dataset with 100s person ids for evaluations. Why not use other state-of-the-art datasets, such as CUHK-PEDES in [1]? Furthermore, in this paper, for Torso and Leg attribute prediction, the training person samples collected from adjacent frames will include lacks of variations although some simple methods, e.g., flipping are used for data augmentations. Why not use some large scale attribute datasets [2] for a pre-trian?
4. The evaluation metric average IoU is also not very convincing to test the performance of person retrieval. As the IoU is calculated over all frames in a video sequence, the quality of tracking algorithm may have a large effects on the final results. For the systemic evaluation of person retrieval, the author

can refer a new recent work in [3].

5. Some typos, “However, Current, trend mainly focuses”, and what is the meaning of “%w IoU >0.4” which is a strange measurement. Is it the mAP score when the threshold of IoU is 0.4?

[1] S. Li, T. Xiao, H. Li, B. Zhou, D. Yue, and X. Wang. Person search with natural language description. In IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[2] D. Li, et al., A richly annotated pedestrian dataset for person retrieval in real surveillance scenarios, In IEEE T-IP 2019.

[3] D. Li, et al. ISEE: An Intelligent Scene Exploration and Evaluation Platform for Large-Scale Visual Surveillance, IEEE T-PDS, 2019.

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.

The novelty of the paper is very limited. And the evaluations are a limited dataset, which is not convincing

6. REVISION OPTION - QUESTION ONE If not accepted in this first round, should the paper be invited to resubmit a revised version of the paper? It would then be considered by the same reviewers and area chair.

REJECT Preferred. Authors will need to submit a SUBSTANTIALLY IMPROVED version for review.

7. REVISION OPTION - QUESTION TWO Please clearly indicate the issues that you feel should be addressed in a revised version such that the paper could potentially be considered for acceptance.

1. Improve the technical novelty

2. Perform more extensive evaluations with large-scale dataset

Reviewer #4

Questions

1. PAPER SUMMARY What is the paper about? Please, be concise (2 to 3 sentences).

This paper studies the problem of person retrieval in surveillance videos using a given semantic description. The authors proposed to use one model for person detection and segmentation and another one for biometric classification. The method outputs bounding box for the person that is most similar to the given description.

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).

This paper proposes a method for person retrieval using semantic description. This is an interesting and important problem that rises in many real-world applications. The proposed approach also seems to be practical and to have good results.

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 presentation of proposed method and results need to be significantly improved. For the most part, the proposed approach is not clearly presented. For instance, the algorithm for "Height Estimation" is not clear and not concretely stated. Also, figures need improvement. Some figures are too large (e.g., figure 5), others not clear (e.g., figure 4). The text in figure 1 is also not readable.

The writing of paper needs major revision. There are too many grammatical errors to report here.

About related work:

The authors need to provide a better discussion on the exact contribution of the paper with respect to related work. There are some statements that vaguely address this. Here, I quote a couple of such statements

"There are some other works, which tries to search a person with other modality of data, such as ..., which are more similar to the problem we aim to tackle in this task." (The authors did not continue to elaborate what are exact differences of proposed method and existing approaches.)

"There are other methods which are more or less similar to our work ...". (The term "more or less" specifically does not help in understanding the exact shortcomings in previous methods that the authors tried to address.)

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.

The problem studied in this paper is valid and important. The paper, however, fails to clearly address the contribution with respect to related work and does not present the method clearly. Also, the language and organization of paper need major revision.

6. REVISION OPTION - QUESTION ONE If not accepted in this first round, should the paper be invited to resubmit a revised version of the paper? It would then be considered by the same reviewers and area chair.

REJECT Preferred. Authors will need to submit a SUBSTANTIALLY IMPROVED version for review.

7. REVISION OPTION - QUESTION TWO Please clearly indicate the issues that you feel should be addressed in a revised version such that the paper could potentially be considered for acceptance.

The organization and language of paper need substantial improvement. The paper, in its current form, is not ready to be published as a conference paper.