

View Reviews

Paper ID

6828

Paper Title

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

Reviewer #1

Questions

1. Summary. In 3-5 sentences, describe the key ideas, experiments, and their significance.

This paper proposes to solve person retrieval in surveillance videos using semantic descriptions. The proposed method utilizes instance segmentation methods to detect pedestrian and then use predefined attributes by constricted order to do pedestrian retrieval. The experiments is good.

2. What are the strengths of the paper? Clearly explain why these aspects of the paper are valuable.

Pedestrian retrieval with pedestrian attribute descriptions is indeed an important problem, which has attracted more and more researchers recently. The experimental results in this paper are promising.

3. What are the weaknesses of the paper? Clearly explain why these aspects of the paper are weak.

Comments:

- 1) Lack of novelty. There is no much insights in this paper. The pipeline to use multiple attributes in predefined order to retrieval pedestrian is common. The author should also try some other attribute order and explain why utilize the given order in the paper.
- 2) The metric to evaluate the pedestrian retrieval performance with IOU is a little confused. As a retrieval problem, the typical metric is mAP, rather than the IOU in detection.
- 3) The dataset descriptions in Section 4.1 are very confused. It is hard to know how many person identities/images/attributes in the training, validation and test stage.
- 4) The author should describe objective function and optimization process in each stages.
- 5) The content in Figure 2 are not clear. Please enlarge the content, e.g. "a,b,c,d".
- 6) The paper writing is not good enough for publication. There are so many grammar errors. Some problems are listed as follow.
 - > "and 0.792 IoU > 0.4%" in Line 130. The representation is hard to understand.
 - > "Figure 2:" in Line 423 should be Figure 3.
 - > "In every video sequence in the training set is labeled with a set of semantic descriptions to describe the person(s) identity." In Line 472. Please deleted "In".
 - > "current state-of-the-art method" -> "current state-of-the-art methods" "Form"->"From" in Line 581.

4. Paper rating (pre-rebuttal)

Weak reject

5. Justification of rating. What are the most important factors for your overall recommendation?

Lack of novelty and the paper writing is not good.

10. Final recommendation based on ALL the reviews, rebuttal, and discussion (post-rebuttal)

Weak reject

11. Final justification (post-rebuttal)

As stated in the comments, I give the weak reject.

Reviewer #2

Questions

1. Summary. In 3-5 sentences, describe the key ideas, experiments, and their significance.

This paper addresses a problem of person retrieval in video surveillance with a semantic description. Semantic description is defined as a set of attributes that later are used in cascade filtering to provide retrieval results. The results on small dataset seem promising, but they are not representative (e.g. re-identification evaluation on only 5 identities/sequences).

2. What are the strengths of the paper? Clearly explain why these aspects of the paper are valuable.

- addressing the person re-id problem as a semantic search is undiscovered area in re-id

3. What are the weaknesses of the paper? Clearly explain why these aspects of the paper are weak.

- minor scientific contribution

- "hand-crafted" set of attributes used as a set of binary filters

- not representative experiments without details on training

- lack of large scale dataset evaluation

4. Paper rating (pre-rebuttal)

Strong reject

5. Justification of rating. What are the most important factors for your overall recommendation?

The paper presents the method that uses Mack RCNN for extracting segmented image of the person and runs multiple heuristic-driven methods in a cascade way to perform retrieval (e.g. head, torso and leg area detection is driven by hard-coded thresholds). Further, the paper proposes to train filters for each attribute independently (where attributes have been ad-hoc defined). Further, these attributes are used in yes/no manner to provide filtering.

Hand crafted set of attributes, as well as the lack of handling missing attributes (e.g. what if classifier makes mistake and incorrectly estimates some set of attributes) make this framework not-scalable.

For training some of attributes (eg. age), DenseNet-161 was used on a very small dataset (e.g. 151 identities). Overall, experiments are performed on a single, very small dataset and some of evaluations are performed on very small subsets (e.g. 5 identities). Please note that recent re-id papers (especially the one that use CNN-based approaches) perform evaluation on one/two magnitude larger datasets (see Market1501, CUHK, DukeMTMC-reID).

10. Final recommendation based on ALL the reviews, rebuttal, and discussion (post-rebuttal)

Strong reject

11. Final justification (post-rebuttal)

No rebuttal has been submitted so I keep the initial ranking.

Reviewer #3

Questions

1. Summary. In 3-5 sentences, describe the key ideas, experiments, and their significance.

This paper studies the problem of person retrieval in video surveillance with a semantic description. The authors designed a cascade filtering strategy for this problem. Overall, the paper proposed a very simple method. The idea is a bit common. The research novelty is limited.

2. What are the strengths of the paper? Clearly explain why these aspects of the paper are valuable.

The authors designed a system for practical application.

3. What are the weaknesses of the paper? Clearly explain why these aspects of the paper are weak.

- 1) There are a lot of tricks in the proposed method.
- 2) The evaluation metric does not make sense. IoU can't represent a certain person being retrieved or not.
- 3) Only one dataset was used to evaluate the proposed method.

4. Paper rating (pre-rebuttal)

Weak reject

5. Justification of rating. What are the most important factors for your overall recommendation?

The novelty is limited.

6. Additional comments.

- 1) A lot of tricks are not clear, for example, what colors were used? why the test person identities were separated into four groups?
- 2) The font of Figure 4 is a little small.
- 3) The reason that choosing a cascade filtering strategy was not illustrated. Why height first, then the torso, leg and finally gender?

10. Final recommendation based on ALL the reviews, rebuttal, and discussion (post-rebuttal)

Weak reject

11. Final justification (post-rebuttal)

No rebuttal was submitted, so I would like to keep my rating.

