



Vandit Jyotindra Gajjar &lt;a1779153@student.adelaide.edu.au&gt;

---

## Congratulations your paper was accepted for the HADCV'21 workshop

---

Microsoft CMT <email@msr-cmt.org>  
Reply-To: Afzal Godil <afzal.godil@nist.gov>  
To: Vandit Gajjar <vanditjyotindra.gajjar@student.adelaide.edu.au>

23 November 2020 at 03:57

Sorry, please disregard my previous email. I had made a mistake in the CMT template. -Afzal

Hi Vandit Gajjar,

Congratulations your paper "PeR-ViS: Person Retrieval in Video Surveillance using Semantic Description" was accepted for the HADCV'21 workshop. The final camera-ready version is due on Nov 27, 2020.

Below are the reviewer's comments for your paper, please incorporate the reviewers' recommendations as time permits, particularly those that are editorial and related to references without requiring additional experiments.

Looking forward to meeting you at the HADCV'21 workshop.

Best regards,

-Afzal for the workshop organizers

Reviewer #1

Questions

### 1. PAPER SUMMARY

This paper proposed a semantics-based query method for person retrieval in video surveillance which has some advantages over traditional image-based query methods.

### 2. PAPER STRENGTHS & WEAKNESSES

Strengths:

1. Semantic information is increasingly used in the field of 3D object retrieval, it is a good attempt for the author to apply person retrieval in video surveillance.
2. Person retrieval in video surveillance is very practical, for example, it is very useful in search notices.

Weaknesses:

1. If there are two people appearing in the same video at the same time, and both of them are male (or female), they are about the same height, body shape and wear similar clothes, then how could you figure out which one is the one you wanted?
2. The paper mentions that image-based query methods have limitations, e.g., no image queries provided, while the proposed semantics-based query method is more general to address this issue. If image queries are provided, can you prove that the semantic query method has higher accuracy than an image-based one?
3. Is it better to take face recognition into consideration? Because if there are two people who look similar or having the same semantic description, face recognition may be able to differentiate them.

### 3. FINAL RECOMMENDATION

Accept

Reviewer #2  
Not Submitted

Reviewer #3  
Questions

#### 1. PAPER SUMMARY

The paper proposes a technique to retrieve persons in video surveillance based on the semantic description of them such as Height; Torso Cloth color, type, and pattern; Leg Cloth color and pattern; and Gender where an image of the person is not always available in real-world scenario. The major contribution is the cascade filtering approach where sequence of those description detections can illuminate a lot of false positives and reduce the search space.

I highly recommend the authors to correct and consider the weak points in my review before submitting any final manuscript.

#### 2. PAPER STRENGTHS & WEAKNESSES

Strengths:

- using the dataset authors claimed achieving new SOTA record compared to previous approaches.
- The filtering approach is interesting way to reduce search space of the query.

Weaknesses:

- The provided plots (specially fig. 4) is of very low quality and hard to look at easily. Even after magnification, the colors of the legend does NOT match the curves!!
- the captions on the workflow of the system is also of very small text font size and not readable!!
- Page 6 from lines 677 to 689 is hard to follow and need to be re-written carefully and in logical order. (e.g. "The other two video sequences" --> This is not clear at all what the authors are talking about!!).
- While the ablation study only talked about the NN architecture, the paper didn't report any other studies on the effect of the semantic descriptors to get an idea which descriptor contributes the most, etc

#### 3. FINAL RECOMMENDATION

Accept

Reviewer #4  
Questions

#### 1. PAPER SUMMARY

To retrieve persons in surveillance video, this paper introduces semantic descriptions, which are known as soft-biometrics, including age, gender, height, cloth type and pattern, etc. It mainly includes two steps: an instance segmentation based on Mask-RCNN is first used to extract pedestrians and corresponding attributes, and then four cascade filters are applied to gradually refine the candidate. The proposed method achieves 0.566 Average IoU, and outperforms the state-of-the-art methods on SoftBioSearch dataset.

#### 2. PAPER STRENGTHS & WEAKNESSES

STRENGTHS: Via combining [4] and [12] effectively, the proposed algorithm obtains the state-of-the-art performance.

WEAKNESSES: Main modules such as Mask-RCNN, cascade filters and DenseNet-161 are presented in [4] and [12] respectively, thereby the novelty of this paper is limited. Moreover, there are other problems:

1. Many writing errors appear in the manuscript such as in line121, line131 and line186 etc, thus the original version should be carefully revised.
2. Figure 2 and Figure 4 with text are quite small and indistinct.
3. The supplementary material mentioned in Figure 7 is missing.

#### 3. FINAL RECOMMENDATION

Borderline

Reviewer #5  
Not Submitted

Reviewer #6

## Questions

## 1. PAPER SUMMARY

The paper proposes a method for person re-identification based on semantic information in form of soft biometrics. The proposed method detects the persons in frames using Mask R-CNN. The soft biometrics are extracted from the detected persons by applying various well-known methods. The soft biometrics classifications are done using Densenet.

## 2. PAPER STRENGTHS &amp; WEAKNESSES

## Strength:

1. The idea of using soft biometrics in person re-identification is an interesting idea.
2. The work seems to be at incremental stage, but seems promising.

## Weakness:

1. Novelty is less. The paper seems to be a mere combination of several existing methods. Mask R-CNN for person detection, different well-known techniques to extract soft-biometrics (height, color, etc.) and Densenet for classification; all are well-known techniques. Novelty is not significant.
2. Since the paper is based on soft-biometric based person re-identifications, a few such techniques should be discussed in the related works section, to establish the motivation behind the proposed approach.
3. IoU may be a good metric to test detection/ segmentation methods. However, for person re-identification problems, this may not be a good metric to validate.
4. The proposed method should be compared with more recent methods for person re-identification. Especially, soft biometric based recent techniques should be compared.

## 3. FINAL RECOMMENDATION

Reject

[Quoted text hidden]