

View Reviews

Paper ID

3

Paper Title

Person Retrieval in Surveillance Video using Height, Color and Gender

Reviewer #1

Questions

1. Provide an overview of the paper

This paper attempts Task 2 in the AVSS Challenge II.

In their work they use a linear filtering approach to soft biometric labelling where each trait can be considered a filter, which filters out detected subjects.

Their work employs some form of Mask R-CNN to both detect and segment pedestrians in the image.

Once pedestrians are detected and segmented the first step detects the height using the Tsai method of calibration which was included as part of the released data.

If the subject meets the height requirements the attributes are classified in turn. First the segmented image is heuristically split into torso and leg regions.

These regions, are then used to classify the colour using the supplied colour snippets for training and AlexNet. Gender is also classified using AlexNet.

2. What are the strengths of the paper?

Their manuscript uses a segmentation and detection module in one, which has the potential to improve attribute classification in future steps.

The use of deep nets for this task is an interesting step forward in the process, and particularly the linear use of soft biometric traits.

3. What are the weaknesses of the paper?

While the linear use of attribute classification is a strength it is also a potential weakness. What happens if the desired subject is not in the frame but there is a subject that meets say the initial 2 criterion? Or what happens if there are errors in the classification process.

4. Suggestions for improvement

Some grammatical issues to be rectified.

line 133: "... interest using description."

line 197: "... (e.g., " -> remove the comma after the e.g.

line 213: "This results in to additional rows ..."

line 222: "Person's height is view and ..."

line 224: "The use of color [15] has the following advantages: " -> the citation seems out of place - please expand it's relevance.

line 349: "... patch is then passing to fine-tuned ..."

line 533: "Thus, algorithm retrieve ..." -> this whole sentence is confusing.

Other issues

line 146: "A study [2] identified 13 soft biometric traits." -> please expand this point as it's unclear what's being said about this study?

Questions?

Was Mask R-CNN off the shelf or was it fine tuned on some data? Please explain either point.

How accurate was Mask R-CNN at actually detecting all pedestrians in the image? False detections? Missed detections? Any idea?

Did you validate either of your AlexNets? Colour? Gender?

Please explain the issues with your methods in your conclusion without referring to figures, actually briefly explain what the issue was here.

5. Recommendation

Accept

Reviewer #2

Questions

1. Provide an overview of the paper

The paper proposes an approach that uses mask rcnn followed by classifiers for gender, height and colour. Gender and colour are classified using a DCNN, while height is obtained using the camera calibration. The traits are applied in a cascade to locate the person.

2. What are the strengths of the paper?

- the approach is simple and fairly well explained. It should be possible for someone else to recreate the system from the paper.

3. What are the weaknesses of the paper?

- cascaded use of classifiers seems to have potential weaknesses if an early classifier makes an error

4. Suggestions for improvement

Parts of the paper are lacking in detail, and further discussion of these points would greatly improve the paper. The following should be addressed:

- A heuristic approach for torso colour is used. Discussion of implications would be good as well as discussion as to why this and not a more automated (but complex) method such as semantic segmentation was not used

- Given that classifiers are applied in a cascade (following Figure 1) it would be interesting to know the accuracy of the individual traits.

- For the person search, is any tracking used, or is it full re-detection each image? Further comments on this would be interesting

- It would be interesting to see outputs of Mask-RCNN visualised with the data, in particular for the negative examples. This would help understand why the approach failed.

- The paper needs to be carefully proof read, in particular the later sections (the earlier sections read pretty well, it's more the back half that needs to be checked.)

5. Recommendation

Borderline