Unsupervised Person Localization In WISAR

Seminar Project of Team B0 | WS21

PHILIPP EBERSTALLER, k11824945, Austria
DOMINIK HEINDL, k11709205, Austria
CARSON WITTWER, k12133590, United States of America

In this project we used an unsupervised algorithm to find occluded, moving people in the forest. We integrated different camera views of a large camera array together and performed anomaly detection in the resulting RGB images. By using different timesteps we were able to confidently detect moving objects in the strongly occluded environment. With our approach of anomaly detection combined with post-processing of said anomalies together with the available temporal data we were able to achieve an average precision (AP) of 57.7% on the available validation set of eleven images.

1 OUR WORK

Image Integration. In order to gather all the information from our array of cameras into a source for processing and anomaly detection, we need to combine these images. First we use the provided homographies in order to align all the images to the central camera point. Once the images are warped correctly, they are merged using a simple technique of averaging the pixel values at each location in order to get our composite of all camera sources. In our testing the mean approach proved to be more useful than the median.

Additionally, as part of the integration process, we identify a mask that contains the locations where a pixel contains color from every input picture, excluding areas where, for example, only 4 images overlap. This allows us to perform future steps on the area of full image overlap and drastically speeds up the computations.

Pre-processing. We calculate a background image, which is a blurred image of the average of all integrated images. We then subtract about 40% of this background from the images to make the anomalies stand out more. (Subtracting 100% of the background image would lead to a lot of noise and in our testing 40% seemed to work the best.)

Anomaly Detection. Our anomaly detection is based on the first main block of the method introduced by [Davy et al. 2018]. We compute a residual image by removing self-similar image components of the original image. This is realised dividing the original image into small patches, 5×5 in our case with a stride of 3. These small patches are then used as input for a k-nearest-neighbour classifier to find similar patches outside of a small square region around the original patch. This ensures that no anomalies with any sort of internal structure are mistaken to be part of the image. After the n most similar patches are found, they are averaged to get the self-similar estimate. The resulting residual image then only contains noise. Other than [Davy et al. 2018] we already use this residual image as a basis for our anomaly detection process rather than performing a statistical test on it.

Anomaly Mask Post-Processing. The resulting anomaly masks have pixel values of about 0 to 15 and still contain a lot of noise. To remove this noise we first apply a Gaussian blur to the masks and then set values smaller than four to 0. To ensure that the anomaly is moving and to ignore anomalies that just occur in individual images, we then calculate the pairwise differences of the anomaly masks between all the time-steps. We then take the average of all those differences, slightly blur them again and use the 0.03% of pixels (about 300) with the highest values as our final mask.

Creating Bounding Boxes. We create the bounding boxes in two steps. First we create contours around all islands of pixels in the anomaly mask and immediately ignore all contours with an area less than 15 pixels. We then combine contours that are closer than 20 pixels from each other and create a rectangle around them.

In the second step we look at each of those rectangles individually. We first calculate a new center for the bounding box which has just the average coordinates of all pixels within the rectangle detected by the anomaly mask. We also calculate the average horizontal and vertical distance of the detected pixels from this new center. To create the bounding box we just take the new center and add the average horizontal or vertical distance in each direction. Since we noticed that our bounding boxes are much smaller than the targets, we also added 20 pixels in each direction to get our final bounding boxes.

2 RESULTS & DISCUSSION

After completion of the process, our system outputs a series of bounding boxes that we believe coordinate with moving humans in the images. Using the provided evaluation utility, we scored an AP of 57.7% of the intersection of union for the validation data set.

Our process of identifying humans is clear on its objective of being a fully unsupervised system. Our process requires no training data at all, as it creates an anomaly mask solely on the content of the current image. Despite the simplicity we believe to have delivered a fairly acceptable result on this challenge. We already made some optimizations to reduce the computational complexity, however the algorithm has a runtime complexity of $O(n^3)$, this means it is extremely sensitive to the input dimensions of the image. With cropping of unwanted regions we were able to keep the runtime in check, but we still see that this has the most room for improvement. Another downside of this approach has to do with the challenge setup. We are not sure how well our algorithm can generalize as most of the used hyperparameters were found by trial and error. For instance many cutoff values were chosen by pure gut feeling and worked out great in this specific case. But we are not sure if this will work as good on another set of images or in any other environment. The last shortcoming is that our algorithm is not able to detect more than one moving object in some of the example images of the validation set.

REFERENCES

Axel Davy, Thibaud Ehret, Jean Michel Morel, and Mauricio Delbracio. 2018. Reducing Anomaly Detection in Images to Detection in Noise. *Proceedings - International Conference on Image Processing, ICIP* (aug 2018), 1058–1062. https://doi.org/10.1109/ICIP.2018.8451059 arXiv:1904.11276