Wheelchair-user Detection Combined with Parts-based Tracking

Ukyo Tanikawa[1], Yasutomo Kawanishi[1], Daisuke Deguchi[2], Ichiro Ide[1],
Hiroshi Murase[1] and Ryo Kawai[3]
[1]*Graduate School of Information Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, Japan*
[2]*Information Strategy Office, Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, Japan*
[3]*NEC Corporation, 1753 Shimo Numabe, Nakahara-ku, Kawasaki-shi, Kanagawa, Japan*

Keywords: Object Detection, Wheelchair User, Crowded Scene, Parts-based Tracking.

Abstract: In recent years, there has been an increasing demand for automatic wheelchair-user detection from a surveillance video to support wheelchair users. However, it is difficult to detect them due to occlusions by surrounding pedestrians in a crowded scene. In this paper, we propose a detection method of wheelchair users robust to such occlusions. Concretely, in case the detector cannot a detect wheelchair user, the proposed method estimates his/her location by parts-based tracking based on parts relationship through time. This makes it possible to detect occluded wheelchair users even though he/she is heavily occluded. As a result of an experiment, the detection of wheelchair users with the proposed method achieved the highest accuracy in crowded scenes, compared with comparative methods.

1 INTRODUCTION

In recent years, various efforts are being made to realize a symbiotic society where people with disabilities can enjoy their lives actively. For example, many public facilities have become handicapped-accessible to support wheelchair users. However, there are still many scenes where they need help from others. In such cases, to provide appropriate support as needed, there has been an increasing demand for a system to detect wheelchair users automatically from surveillance video.

However, in an actual environment such as railway stations, many pedestrians often surround wheelchair users. Figure 1 shows an example of a wheelchair user moving in a crowded scene. In a crowded scene like this, there is a problem that detection often fails since the whole body of a wheelchair user is not visible because of occlusions caused by surrounding pedestrians. In this paper, we aim to detect occluded wheelchair users in a crowded scene, and propose a detection method robust to occlusions.

A detector often suffers when the target is heavily occluded. Tracking based on their past positions enables us to locate occluded targets even when the detector cannot detect them. However, when the tracking target is occluded, the tracking accuracy declines.

Figure 1: Example of a wheelchair user moving in a crowded scene. Occlusions by surrounding pedestrians are often observed.

We have observed that some parts of a wheelchair user are visible even if his/her body is almost occluded, because the width and the depth of wheelchair users are larger than those of pedestrians in general. If the parts are visible, we can roughly estimate his/her bounding box. In this paper, based on the observation, we propose a method which combines detection by a detector with parts-based tracking. When a traditional detector could not detect a target due to occlusions, the proposed method can estimate its location based on parts-tracking results.

Since the size of the parts is small in general, it is difficult to distinguish them from other objects. Hence, a parts tracker which only considers their appearance can drift easily. To reduce the drift, we in-

Tanikawa U., Kawanishi Y., Deguchi D., Ide I., Murase H. and Kawai R.
Wheelchair-user Detection Combined with Parts-based Tracking.
DOI: 10.5220/0006101101650172
In *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017)*, pages 165-172
ISBN: 978-989-758-226-4

troduce part tracking confidence and parts relationship through time; The proposed method calculates the tracking confidence of each part of a target. The parts with high confidence are tracked based on their appearances. The positions of the parts with low confidence are predicted based on their past trajectories and inter-parts positional relationships.

In summary, our contributions include the proposal of:

- A framework of wheelchair-user detection robust to occlusions by combining a detector with parts-based tracking.
- A parts-based tracking method which considers trajectories and inter-parts positional relationships to predict positions of parts with low confidence.

The rest of the paper refers to related works in Section 2, describes the proposed framework in Section 3 and the proposed parts-based tracking method in Section 4, reports evaluation results in Section 5, and concludes the paper in Section 6.

2 RELATED WORKS

Dalal and Triggs proposed an object detection method using Histogram of Oriented Gradients (HOG) features (Dalal and Triggs, 2005). HOG is a feature descriptor robust to local shape deformations, illumination variations, and effects of shades. However, HOG cannot handle large pose deformations. In contrast, Felzenszwalb et al. proposed an object detection method using Deformable Part Model (DPM), which represents an object model with a set of parts (Felzenszwalb et al., 2010). DPM is robust to pose deformations by considering fine shape and position of each part. The position is treated as latent variables and automatically learned by using Latent SVM (Felzenszwalb et al., 2010). However, DPM has a problem that its detection accuracy degrades when the parts are occluded.

Myles et al. proposed a detection method specialized for wheelchair users based on the detection of wheels and faces of their users (Myles et al., 2002). In this method, wheels of wheelchairs are detected by using the Hough transform, and faces of their users are detected based on their color features. Then, their 3-D poses are constructed by 2-D ellipse projection. However, this method needs accurate calibration in advance. Huang et al. proposed a method of wheelchair-user detection from a single camera with no calibration (Huang et al., 2010). This method uses HOG and Contrast Context Histogram features (Huang et al., 2006), and a hierarchical cascade classifier using AdaBoost is built. However, this method does not consider occlusions of wheelchair users, so in a crowded scene, it cannot detect them accurately.

Henriques et al. proposed a method for single object tracking using Kernelized Correlation Filter (KCF) tracker (Henriques et al., 2015). KCF tracker achieves good performance with high speed. It is a method based on kernel ridge regression, and is a kind of correlation-filter-based tracking methods (Bolme et al., 2009; Bolme et al., 2010). Correlation-filter-based trackers can calculate tracking confidence using the Peak-to-Sidelobe Ratio (PSR), which quantifies the strength of correlation peak relative to an area around the peak in a response map (Bolme et al., 2010).

Zhang et al. proposed a method for multi-person tracking combining person detection by a detector with visual object tracking (Zhang et al., 2012). This method represents target appearance with a set of templates gathered from detections, and tracking is performed by alternating mean-shift tracking and Kalman filtering. This enables an estimation of their location even if the detector cannot detect them. However, this method does not take into account occlusions of tracking targets, so in a crowded scene, it cannot track them accurately.

There are tracking methods which explicitly consider the target's partial occlusions. Pan and Hu proposed a tracking method which handles occlusions by exploiting spatio-temporal context information (Pan and Hu, 2007). However, this method does not consider heavy occlusions.

In summary, these conventional methods cannot handle heavy occlusions of wheelchair users well.

3 FRAMEWORK OF WHEELCHAIR USERS DETECTION COMBINED WITH PARTS-BASED TRACKING

As mentioned above, detection of wheelchair users in a crowded scene is challenging due to heavy occlusions. These occlusions are often caused by surrounding pedestrians. Since it is difficult to detect occluded targets from only a single frame, we introduce a framework with parts-based tracking across multiple frames, which is introduced in Section 4.

Figure 2 shows the process flow of the proposed framework. In the training phase, a wheelchair-user detector is trained. In the detection phase, wheelchair users are detected from each frame of an input sequence by using the trained detector. The detections

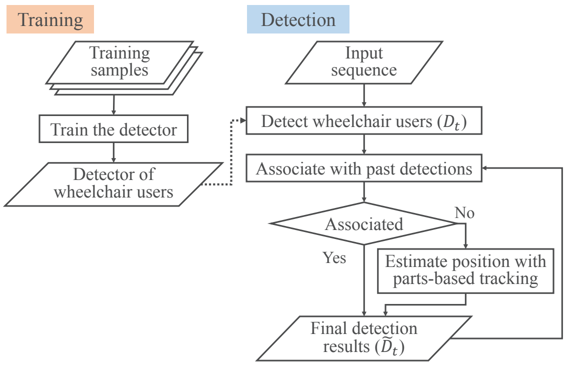


Figure 2: Process flow of the proposed framework.

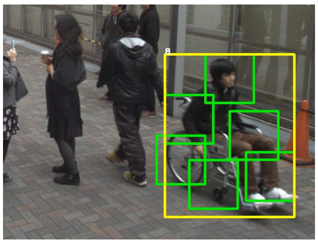


Figure 3: Example of detection using DPM.

from consecutive frames are associated to construct their trajectories. When some detections were not associated, the proposed parts-based tracking is performed to estimate their locations.

3.1 Detection by a Parts-based Detector

Full bodies of wheelchair users and their parts in each frame of an input sequence are detected by a parts-based detector. In this paper, we use DPM (Felzenszwalb et al., 2010) which can simultaneously detect both of them. Figure 3 shows an example of detections using DPM for wheelchair users.

In the training phase, a DPM detector for wheelchair users is trained with many positive and negative images. In the detection phase, wheelchair users are detected from each frame of input sequences by the trained DPM detector.

3.2 Association of the Detections

For each frame of an input sequence, detection results of wheelchair users are associated to construct their trajectories. Let $\tilde{D}_{t-1} = \{d_{t-1}^{(1)}, d_{t-1}^{(2)}, \ldots, d_{t-1}^{(n_{t-1})}\}$ be the final detection results obtained with the proposed method in the $(t-1)$-th frame, and $D_t = \{d_t^{(1)}, d_t^{(2)}, \ldots, d_t^{(n_t)}\}$ be the detection results obtained with the parts-based detector in the t-th frame. First, the similarity between each pair in $\tilde{D}_{t-1}$ and D_t is calculated to find similar detection results. In this paper,

Figure 4: Example of the detection association process.

we use an overlap ratio $\Omega(d_{t-1}^{(i)}, d_t^{(j)})$ between the pair of detections $(d_{t-1}^{(i)}, d_t^{(j)})$ as the similarity, which is defined as follows:

$$\Omega(d_{t-1}^{(i)}, d_t^{(j)}) = \frac{|d_{t-1}^{(i)} \cap d_t^{(j)}|}{|d_{t-1}^{(i)} \cup d_t^{(j)}|}. \tag{1}$$

The similarity $S(d_{t-1}^{(i)}, d_t^{(j)})$ between $d_{t-1}^{(i)}$ and $d_t^{(j)}$ is defined as follows:

$$S(d_{t-1}^{(i)}, d_t^{(j)}) = \begin{cases} \Omega(d_{t-1}^{(i)}, d_t^{(j)}) & \text{if } \Omega(d_{t-1}^{(i)}, d_t^{(j)}) > \theta_\Omega \\ 0 & \text{otherwise} \end{cases}. \tag{2}$$

Detection results are associated by selecting the pair of detections which has a larger similarity than a threshold. Figure 4 shows an example of association of the detection results.

3.3 Estimation using Parts-based Tracking

When the detector lost the target detected more than θ_d times continuously due to occlusions, his/her position is estimated by tracking. While his/her fullbody-tracking is difficult due to occlusions, some parts of the body are often visible even if it is almost occluded. We perform the parts-based tracking introduced in Section 4 to estimate his/her positon. Detection results of parts by the parts-based detector are utilized as an initial bounding box of parts-tracking.

Parts-based tracking is conducted based on the position of the target in the $(t-1)$-th frame, and its position after the t-th frame is estimated from its past trajectory and the position of its confidently-tracked parts.

Parts-based tracking continues up to f_1 frames. It terminates in the following cases:

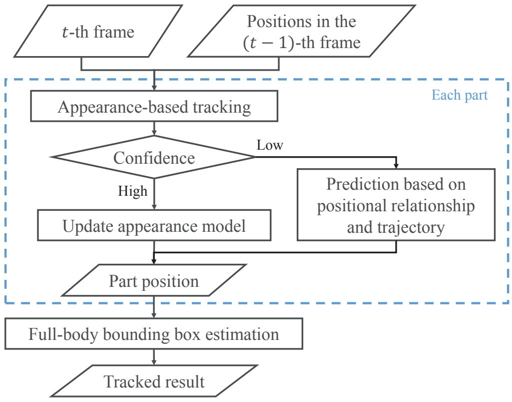


Figure 5: Process flow of the proposed parts-based tracking method.

- The tracked result and the detection result were associated successfully, i.e., the target was detected by the parts-based detector again before f_1 frames passed.
- All parts of the target were occluded for f_2 consecutive frames. This can suppress false detections caused by failed parts-based tracking.

4 PARTS-BASED TRACKING

Parts-based approach is robust to the target's occlusions. Since the size of parts is small, it is difficult to distinguish them from other objects. Hence, parts-based tracking which only considers their appearance can drift easily. To track them accurately, the proposed parts-based tracking method compensates the position of parts considering their past trajectories and inter-parts positional relationships.

Figure 5 shows the process flow of the proposed parts-based tracking method. First, the proposed method tracks each part based on its appearance and calculates each tracking confidence. If the tracking confidence of the part is high, its appearance model is updated. If the confidence is low, its position is extrapolated based on their past trajectories and inter-parts positional relationships. We integrate these information into a score map on the center position of the target, and adopt the position that maximizes this score. In the end, the full-body bounding box is estimated based on the parts locations.

4.1 Appearance-based Tracking and Confidence Calculation

The proposed method tracks each part of a target using KCF tracker (Henriques et al., 2015). KCF tracks a target convolving an input image with a filter designed to produce correlation peaks for the target in a response map, while producing low responses to background. The filter is updated over time to adapt to appearance change.

The proposed method tracks each part based on its appearance and calculates each tracking confidence. When it is difficult to track the part (e.g., its size is small, or it is occluded), its confidence tends to get lower. Therefore, we change the tracking method according to the confidence.

In the following explanation, we describe the process for each part. First, the response map of KCF tracker for the part is caluculated. Next, tracking confidence is calculated from the response map. We utilize the Peak-to-Sidelobe Ratio (PSR) of the response map as the tracking confidence. PSR quantifies the strength of correlation peak relative to the sidelobe area in a response map. In this paper, we define the sidelobe as a square area around the peak which has 15% area of the response map.

The parts which have higher PSR values than a threshold θ_{PSR} are recognized as highly confident. The positions of the parts with high confidence are set to be the positions of correlation peaks in their response map. In contrast, the tracked results which has lower confidence than the threshold are unreliable. We estimate their positions by a method introduced in Section 4.2.

Note that the appearance model of each KCF tracker is updated over time, but updating a lowly confident target's model leads to the decline of tracking accuracy. Therefore, we update models of parts only when they have high confidence.

4.2 Prediction of Parts Positions

The positions of the parts with low confidence are estimated based on their trajectories and inter-parts positional relationship. We integrate these information into score map S on the center position of the part, and adopt the position that maximizes this score. Let $p^{(i)}$ $(i = 1, \ldots, n)$ be the i-th part of the target, P_l be the set of parts with low confidence, and P_h be the set of parts with high confidence in the current frame t. Note that $P_l \cap P_h = \varnothing$, $|P_l \cup P_h| = n$ holds. The center position $\hat{\boldsymbol{x}}_t^{(i)}$ of the low-confident part $p^{(i)} \in P_l$ in the t-th frame is estimated as follows:

$$\hat{\boldsymbol{x}}_t^{(i)} = \arg\max_{\boldsymbol{x}_t^{(i)}} S(\boldsymbol{x}_t^{(i)}), \tag{3}$$

where $\boldsymbol{x}_t^{(i)} = (x_t^{(i)}, y_t^{(i)})$ is the center position of the part $p^{(i)}$ in the image coordinate. $\hat{\boldsymbol{x}}_t^{(i)}$ will be the position

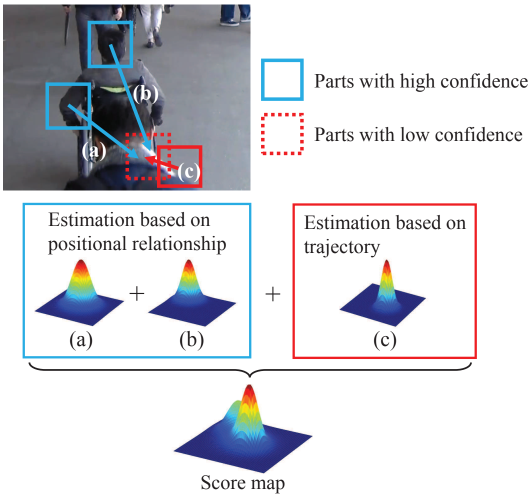


Figure 6: Calculation of the score on the position of the part recognized as occluded.

that maximizes the score $S(\boldsymbol{x}_t^{(i)})$. The width and the height of the estimated bounding box of the part are set to be the same as those in the $(t-1)$-th frame. The score map S on position $\boldsymbol{x}_t^{(i)}$ of part $p^{(i)} \in P_l$ in the t-th frame is defined as follows:

$$S(\boldsymbol{x}_t^{(i)}) = \sum_{p^{(j)} \in P_h} P^b(\boldsymbol{x}_t^{(i)} | \boldsymbol{x}_t^{(j)}, \boldsymbol{x}_{t-1}^{(i)}, \boldsymbol{x}_{t-1}^{(j)}) + \lambda P^u(\boldsymbol{x}_t^{(i)} | \boldsymbol{x}_{t-1}^{(i)}, \boldsymbol{x}_{t-2}^{(i)}). \quad (4)$$

The first term in the right-hand side of Equation (4) is the sum of scores on the position of part $p^{(i)}$ based on inter-parts positional relationships in the $(t-1)$-th frame. The more parts there are with high confidence, the more reliable and larger this score is. The second term in the right-hand side of the Equation (4) is the scores based on its trajectory. λ is the trade-off between the first term and the second term.

The score map P^b based on inter-parts positional relationship between $p^{(j)} \in P_h$ and $p^{(i)} \in P_l$ in the t-th frame is modeled by the sum of bivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}_t^{b,(i)}, \boldsymbol{\Sigma}_t^{b,(i)})$ as shown in Figure 6. The mean vector $\boldsymbol{\mu}_t^{b,(i)}$ and the variance-covariance matrix $\boldsymbol{\Sigma}_t^{b,(i)}$ are defined as follows:

$$\boldsymbol{\mu}_t^{b,(i)} = \boldsymbol{x}_t^{(j)} + (\boldsymbol{x}_{t-1}^{(i)} - \boldsymbol{x}_{t-1}^{(j)}), \quad (5)$$

$$\boldsymbol{\Sigma}_t^{b,(i)} = \begin{pmatrix} \sigma_{x,t,(i)}^2 & 0 \\ 0 & \sigma_{y,t,(i)}^2 \end{pmatrix}. \quad (6)$$

The mean vector $\boldsymbol{\mu}_t^{b,(i)}$ is the sum of the position of $p^{(j)}$ and an offset vector from $p^{(j)}$ to $p^{(i)}$ in the $(t-1)$-th frame. Diagonal components $\sigma_{x,t,(i)}, \sigma_{y,t,(i)}$ of the variance-covariance matrix $\boldsymbol{\Sigma}_t^{b,(i)}$ is calculated as follows:

$$\sigma_{x,t,(i)} = \frac{w_{t-1}}{s} \left(1 - \frac{\mathrm{PSR}\left(p^{(j)}, t\right)}{\sum_{p^{(k)} \in P_h} \mathrm{PSR}\left(p^{(k)}, t\right)} \right), \quad (7)$$

$$\sigma_{y,t,(i)} = \frac{h_{t-1}}{s} \left(1 - \frac{\mathrm{PSR}\left(p^{(j)}, t\right)}{\sum_{p^{(k)} \in P_h} \mathrm{PSR}\left(p^{(k)}, t\right)} \right), \quad (8)$$

where w_{t-1} and h_{t-1} are the width and the height of $p^{(i)}$ in the $(t-1)$-th frame respectively. The larger they are, the shorter and wider the normal distribution becomes. $\mathrm{PSR}(p^{(j)}, t)$ denotes the PSR of $p^{(j)}$ in the t-th frame. The larger $\mathrm{PSR}(p^{(j)}, t)$ relative to that of other parts with high confidence encourages smaller $\sigma_{x,t,(i)}$ and $\sigma_{y,t,(i)}$, i.e., the lower the confidence is, the shorter and wider the distribution becomes. s is a scale parameter.

The score map P^u based on the trajectory of $p^{(i)}$ is also modeled by the bivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}_t^{u,(i)}, \boldsymbol{\Sigma}_t^{u,(i)})$. The mean $\boldsymbol{\mu}_t^{u,(i)}$ of the distribution is defined as follows:

$$\boldsymbol{\mu}_t^{u,(i)} = \boldsymbol{x}_{t-1}^{(i)} + (\boldsymbol{x}_{t-1}^{(i)} - \boldsymbol{x}_{t-2}^{(i)}). \quad (9)$$

$\boldsymbol{\mu}_t^{u,(i)}$ is the sum of the position in the $(t-1)$-th frame and the displacement vector from the $(t-2)$-th frame to the $(t-1)$-th frame. The variance-covariance matrix $\boldsymbol{\Sigma}_t^{u,(i)}$ is a diagonal matrix same as Equation 6, where $\sigma_{x,t,(i)}$ and $\sigma_{y,t,(i)}$ are set to be in proportion to the width and the height of $p^{(i)}$ in the $(t-1)$-th frame, respectively.

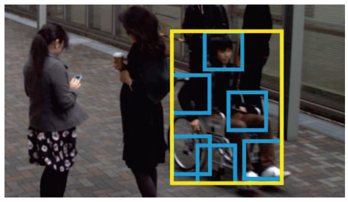


Figure 7: Example of the estimation of the full-body bounding box from part bounding boxes.

4.3 Full-body Bounding Box Estimation

In each frame, the tracked results of parts are put together to estimate a full-body bounding box of the target. The bounding box of the whole target is defined to be a minimum bounding box including all bounding boxes of parts. Figure 7 shows an example of this integration. The inner small rectangles are bounding boxes of parts, and the outer large rectangle is the estimated bounding box of the whole target.

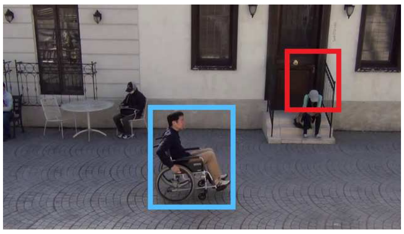


Figure 8: Example of positive and negative training samples.

5 EXPERIMENT

5.1 Experimental Condition

To evaluate the effectiveness of the proposed method in the detection of wheelchair users under a crowded scene, we conducted an experiment. In the experiment, they were detected from video sequences captured in an environment where many pedestrians surrounded them. We compared the following methods:

- `DPM`: Using only DPM detector.
- `DPM + Full-body tracking`: Using DPM detector combined with full-body tracking of targets.
- `DPM + Multi-tracker`: Using DPM detector combined with multi-person tracker (Zhang et al., 2012).
- `DPM + Parts-tracker` (Proposed method): Using DPM detector combined with parts-based tracking.

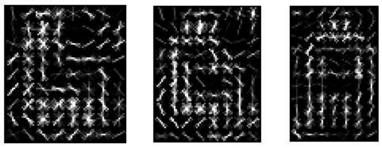
For the evaluation of `DPM + Multi-tracker`, we used the publicly available implementation (Zhang et al., 2013) provided by the authors. We set the parameters for tracking as $f_1 = 20$ frames and $f_2 = 15$ frames.

The overlap ratio between detections by each method and the ground truth was calculated. Detections are considered to be correct when they overlapped more than 50% with the ground-truth bounding box. As an evaluation criterion of detection accuracy, we employed precision, recall, and F-measure.

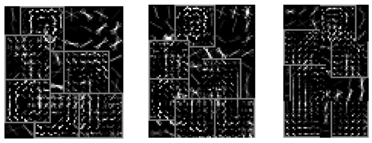
5.2 Datasets

5.2.1 Training Data

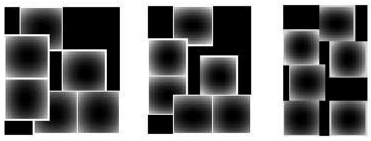
In the experiment, 2,400 images of wheelchair users captured both indoors and outdoors were used as positive samples to train the DPM detector. For each training image, we annotated bounding boxes of wheelchair users manually. As negative samples, we prepared 4,800 images randomly cropped from the background of the training images. Figure 8 shows an example of positive and negative training samples. The larger rectangle in blue shows a positive sample, and the smaller rectangle in red shows a negative sample.



(a) Root filters



(b) Part filters



(c) Deformation costs

Figure 9: Three-components DPM trained for wheelchair users.

5.2.2 Test Data

As test data, we prepared seven video sequences captured outdoors. The size of each frame in the test sequences was $1,280 \times 1,024$ pixels. The length of each sequence was from approximately 30 seconds to 1 minutes, with a frame rate of 6 fps. The number of images in the test sequences was 1,621 and included a cumulative total of 1,175 wheelchair users.

Each frame in the test sequences included at most a single wheelchair user that was often occluded by pedestrians around him/her. There were two cases that wheelchair users existed in the initial frame or entering the frame. Wheelchair users exiting the frame were also included.

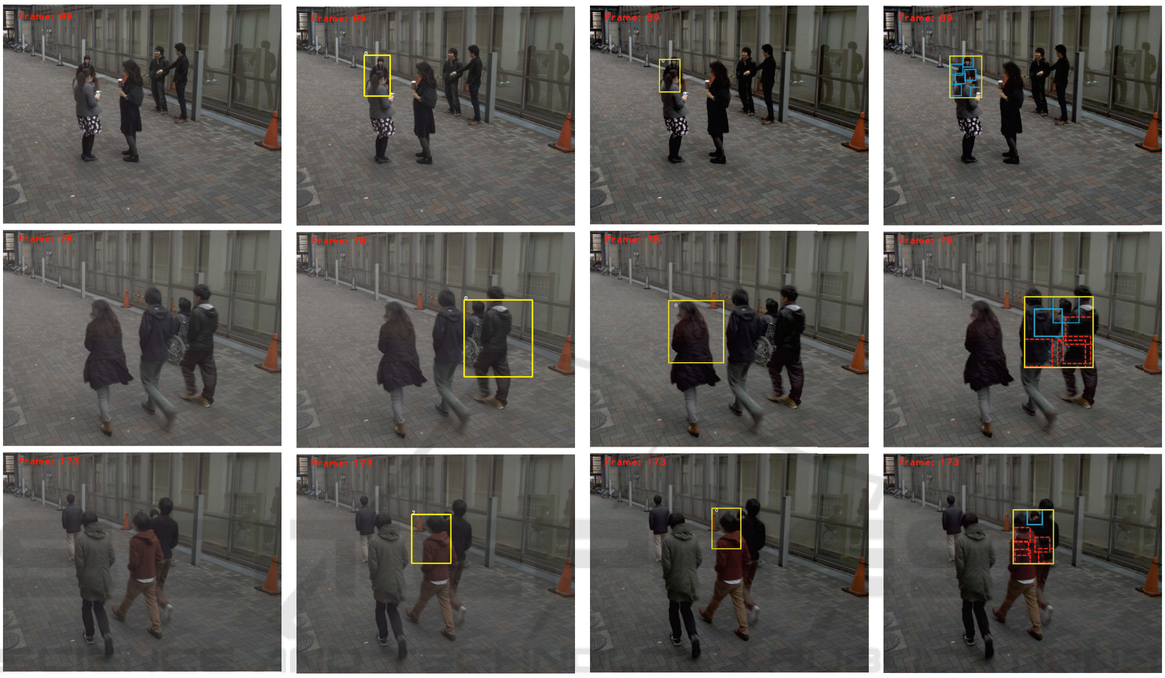
For each frame of the sequences, bounding boxes of wheelchair users were manually annotated as ground-truth for evaluation. In case a wheelchair user was occluded, we annotated a likely bounding box by considering the context.

5.3 Model of Wheelchair Users

In the experiment, we used a three-components DPM detector for detection. In training of the DPM detector, training samples were divided into three clusters based on their aspect ratio, and the model of

Table 1: Detection accuracy of wheelchair users by each method.

	Criterion		
Method	Precision	Recall	F-measure
`DPM`	**0.975**	0.671	0.795
`DPM + Full-body tracking`	0.789	0.808	0.798
`DPM + Multi-tracker` (Zhang et al., 2012)	0.516	0.585	0.548
`DPM + Parts-tracker` (Proposed method)	0.859	**0.848**	**0.853**



(a) `DPM` (b) `DPM + Full-body tracking` (c) `DPM + Multi-tracker` (Zhang et al., 2012) (d) `DPM + Parts-tracker` (Proposed method)

Figure 10: Examples of detections by each method.

wheelchair users was constructed for each cluster. The number of parts each DPM model included was experimentally set to six. The training result for wheelchair users is visualized in Figure 9. In the experiment, the publicly available code of DPM published by Girshick et al. (Girshick et al., 2012) was used.

5.4 Results & Discussions

Table 1 shows the result of detections from test sequences. This result indicates that the proposed method is more accurate on recall and F-measure than other comparative methods. From this result, the effectiveness of the proposed method (`DPM + Parts-tracker`)for detection in a crowded scene can be confirmed. Note that the proposed method is less accurate in precision than `DPM`. This is because the failure of tracking leads to an increase of false positives. However, the proposed method achieved the highest precision of the methods which used tracking. This indicates that the proposed parts-based tracking is more accurate than other tracking methods. The detection accuracy of `DPM + Multi-tracker` is worse than `DPM`. Since `DPM + Multi-tracker` used full-body tracking, the tracking often failed when targets were occluded. In contrast, the proposed method used parts-based tracking, so it improved the detection accuracy even if the targets were occluded.

Figure 10 shows examples of detections by the comparative methods and the proposed method. Each row of Figure 10 shows the detection result by each method in the same frame of the test sequences. In the figure of the proposed method, the inner small rectangles indicated by a broken line are predicted bounding boxes of parts with low confidence. The other inner rectangles are bounding boxes of parts with high confidence. The outer large rectangles are the estimated full-body bounding boxes of the targets. The results in the first row indicate that the target which could not be detected by the DPM detector was successfully detected by being combined with tracking. These results show the effectiveness of combining detection with tracking. Moreover, the results in the second row and the third row show that the proposed method estimated bounding boxes of wheelchair users more accurately than other comparative methods. The proposed parts-based tracking could estimate the bounding boxes even if most of the parts were occluded. These results show that proposed parts tracking is robust against heavy occlusions and it can compensate false negatives of the detector satisfactorily.

6 CONCLUSIONS

In this paper, we proposed a method for detecting wheelchair users accurately in a crowded scene. Detection of wheelchair users was difficult when they were occluded, but the proposed method coped with it by combining the detector with parts-based tracking. To track the parts of wheelchair users accurately, the proposed method estimated the position of parts with low tracking confidence based on their trajectories and inter-parts positional relationships. Experimental results showed that the proposed method can detect them in a crowded scene more accurately than comparative methods.

As future work, we will consider a more effective score function in parts-based tracking to further improve locating of parts with low confidence. We will also modify the method for associating the detection results. In addition, we will introduce sophisticated motion dynamics of wheelchair users.

ACKNOWLEDGEMENTS

Parts of this research were supported by MEXT, Grant-in-Aid for Scientific Research. We would like to thank the members of the laboratory for their cooperation as subjects for creating the dataset.

REFERENCES

Bolme, D. S., Beveridge, J. R., Draper, B., and Lui, Y. M. (2010). Visual object tracking using adaptive correlation filters. In *Proceedings of the 23rd IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2544–2550.

Bolme, D. S., Draper, B. A., and Beveridge, J. R. (2009). Average of synthetic exact filters. In *Proceedings of the 22nd IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2105–2112.

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Proceedings of the 18th IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 886–893.

Felzenszwalb, P. F., Girshick, R. B., McAllester, D., and Ramanan, D. (2010). Object detection with discriminatively trained part based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1627–1645.

Girshick, R. B., Felzenszwalb, P. F., and McAllester, D. (2012). Discriminatively trained deformable part models, release 5. Available at: http://people.cs.uchicago.edu/∼rbg/latent-release5/ [Accessed 14 Sept. 2016].

Henriques, J. F., Caseiro, R., Martins, P., and Batista, J. (2015). High-speed tracking with kernelized correlation filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(3):583–596.

Huang, C.-R., Chen, C.-S., and Chung, P.-C. (2006). Contrast context histogram — A discriminating local descriptor for image matching. In *Proceedings of the 18th IEEE International Conference on Pattern Recognition*, volume 4, pages 53–56.

Huang, C.-R., Chung, P.-C., Lin, K.-W., and Tseng, S.-C. (2010). Wheelchair detection using cascaded decision tree. *IEEE Transactions on Information Technology in Biomedicine*, 14(2):292–300.

Myles, A., Lobo, N. D. V., and Shah, M. (2002). Wheelchair detection in a calibrated environment. In *Proceedings of the 5th Asian Conference on Computer Vision*, pages 706–712.

Pan, J. and Hu, B. (2007). Robust occlusion handling in object tracking. In *Proceedings of the 20th IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1–8.

Zhang, J., Presti, L. L., and Sclaroff, S. (2012). Online multi-person tracking by tracker hierarchy. In *Proceedings of the 9th IEEE International Conference on Advanced Video and Signal-Based Surveillance*, pages 379–385.

Zhang, J., Presti, L. L., and Sclaroff, S. (2013). Online multi-person tracking by tracker hierarchy. Available at: http://cs-people.bu.edu/jmzhang/tracker_hierarchy/Tracker_Hierarchy.htm [Accessed 14 Sept. 2016].