Robust Head-shoulder Detection using Deformable Part-based Models

Enes Dayangac, Christian Wiede, Julia Richter and Gangolf Hirtz
Faculty of Electrical Engineering and Information Technology, Technische Universität Chemnitz, Chemnitz, Germany

Keywords: Person Detection, Head-shoulder Detection, Ambient Assisted Living, Latent SVM, DPM, ACF-Detector.

Abstract: Conventional person detection algorithms lack of robustness, especially when the person is partially occluded. We propose thereby a robust head-shoulder detector in 2-D images using deformable part-based models. This detector can be used in a variety of applications such as people counting and person dwell time measurements. In experiments, we compare the head-shoulder detector with the full body detector quantitatively and analyze the robustness of the detector in realistic scenarios. In the results, we show that the model learned with our method outperforms other methods proposed in related work on an ambient assisted living application.

1 INTRODUCTION

Person detection finds application in a variety of fields such as public security, automotive as well as in the business sector. It furthermore plays a crucial role in the upcoming application field of Ambient Assisted Living (AAL). The notion of AAL is to assist elderly people during their daily life with the help of innovative technology.

In our study, we have designed an AAL system that applies optical sensors. One essential part of this system is a machine learning-based person detection algorithm. This paper focuses on the detection of persons within a complex living environment. In this context, a number of challenges have to be addressed.

Firstly, a person detection algorithm is required to cope with varying illumination conditions and large distances between persons and sensors. Secondly, in contrast to many other studies, a wide-angle camera is mounted on the ceiling of a room, so that a large area of the scene can be covered. As a result, the person's lower part is occluded by the torso, especially when the person is standing very close to the camera. In typical AAL scenarios, such as sitting at a table, the legs are also not visible either. However, the person models used in state-of-the-art algorithms are trained with samples of the full body from frontal view. Consequently, these models show inadequate results and are therefore not suitable for the presented application.

For this reason a head-shoulder model for that special camera set-up has been trained in the presented study. As there are no appropriate public datasets available, we recorded a variety of sequences under different lighting and distance conditions. The performance of the head-shoulder model was analyzed using samples of our own dataset and samples from a public dataset. We have chosen two state-of-the-art algorithms that proved to work effectively with full body models in front-view scenarios and modified them by adapting the head-shoulder model.

2 RELATED WORK

In recent years the number of people detection algorithms has been steadily increasing. In the past, pedestrian detection was pursued by many researchers in the context of Driver Assistance Systems. Dollár et al. presented an overview about state-of-the-art algorithms for pedestrian detection (Dollár et al., 2012). They especially focussed on sliding window techniques, which give promising results for low and medium resolution images in which segmentation and key point based approaches proved to be unsuitable. One of the first sliding window approaches applied a support vector machine that was trained with Haar Wavelets (Papageorgiou and Poggio, 2000). Building upon this concept the approach of Viola and Jones introduced integral images for fast feature calculation (Viola and Jones, 2001). They used AdaBoost as training method and a cascade structure for an efficient detection with a reduced number of false positives. This method serves as a basis for several mod-

Dayangac E., Wiede C., Richter J. and Hirtz G..
Robust Head-shoulder Detection using Deformable Part-based Models.
DOI: 10.5220/0005266002360243
In *Proceedings of the 10th International Conference on Computer Vision Theory and Applications* (VISAPP-2015), pages 236-243
ISBN: 978-989-758-090-1

ern detectors.

The development of advanced, highly-descriptive feature types was essential with regard to the accuracy of detection algorithms. Especially gradient-based features like Scale-Invariant Feature Transform (SIFT) (Lowe, 2004) and Histogram of Oriented Gradients (HOG) (Dalal and Triggs, 2005) played an important role in this context. Nowadays HOG features are the most frequently used descriptors for person detection. A considerable amount of studies, such as (Zhu et al., 2006) and (Shashua et al., 2004), focused on enhancing the detection results. Other studies investigated features that are based on shape (Gavrila, 2007), (Sabzmeydani and Mori, 2007) and motion information (Viola et al., 2003),(Wojek et al., 2009).

Further developments combined several different features in order to get more powerful descriptors. Since HOG is rated as one of the most effective single feature, several studies investigated combinations with HOG and other features. Wojek and Schiele combined HOG, Haar-like features and shape features (Wojek and Schiele, 2008). This approach was later extended by color self similarity and motion information (Walk et al., 2010).

Dollár et al. proposed another method (Dollár et al., 2009a), which takes up the idea of Viola and Jones (Viola and Jones, 2001). They modified the original approach by searching for Haar-like features in different channels, such as the LUV color channel as well as gray-scale, gradient magnitude and orientation images. In one of their following works, they extended this approach to a multi-scale detection (Dollár et al., 2010). In this way processing time was reduced because features can be computed from a nearby scale. Another promising approach are part-based models. In contrast to the previously described approaches, the part-based models are constructed in such a way that not only the person as a whole model is considered, but different parts of the person are described and used for classification. By constructing these part models, variances caused by rotation and occlusion can be reduced. One of the algorithms providing the best results so far is the deformable part-based models approach (DPM). It is a discriminative part-based approach, whereby unknown part positions are described as latent variables (Felzenszwalb et al., 2008), (Felzenszwalb et al., 2010). As they utilized a SVM for classification, this classifier is called Latent SVM. This approach was later extended by Park et al. to a multi-resolution model (Park et al., 2010). Another approach that dealt with partial occlusion was based on an Edgelet detector, which is a kind of shape detector combined with AdaBoost (Wu and Nevatia, 2005).

Beside algorithms processing single, monocular images, there is furthermore a variety of 3-D based approaches. Kirchner et al. segment a person's point cloud into horizontal slices (Kirchner et al., 2012). The span of each slice is accumulated in a feature vector that is used for training a SVM. Richter et al. localize persons in 3-D world coordinates by firstly generating foreground hypotheses on the world z-map and then projecting 3-D points onto a virtual overhead view (Richter et al., 2014). In this study, the authors make the assumption that foreground objects of a certain size are most probable persons. However, it could be sensible to validate the detected persons by means of a 2-D based person detection algorithm.

The latest 2-D person detection algorithms utilize models that are composed of the full body. Since the camera is monitoring the scene at a certain angle with respect to the ceiling in many AAL applications (side and top view if the person is standing near the camera), the full body model is not suitable. In general, head-shoulder detection is more reliable than the full-body detection even in highly occluded cases (Tu et al., 2013). Therefore a new, more efficient model has to be trained. As the shoulder part is always visible in this camera set-up, it is rational to train a model with samples showing the person's shoulder part. There already exist several approaches for head-shoulder detection. Li et al. proposed a method which detects the head-shoulder part by combining a Viola-Jones type classifier and a local HOG feature-based AdaBoost classifier (Li et al., 2009). They furthermore showed how to track head-shoulder parts by a particle filter approach. An attention-based foreground segmentation was combined with a multi-view cascade to detect head-shoulder parts for video surveillance (Tu et al., 2013). Wang et al. introduced a new edge feature called En-Contour (Wang et al., 2013). However, these mentioned methods have not been studied and tested in AAL scenarios.

At present there is no public head-shoulder database available that can be used for training. Commonly-known datasets such as INRIA (Dalal and Triggs, 2005), PASCAL (Everingham et al., 2010), Caltech (Dollár et al., 2009b), ETH (Ess et al., 2007) and Daimler (Enzweiler and Gavrila, 2009) only provide sequences where the camera is installed at approximately the same height as the recorded persons. Therefore the whole body is always visible and the optical axis is almost perpendicular to the person's main axis. For this reason we recorded an own dataset for both training and testing purposes.

3 METHODS

Deformable part-based models (Felzenszwalb et al., 2010) can overcome major challenges arising in object detection. Effects such as deformation, occlusion and viewpoint changes can be managed to a certain extent. However, in many cases in AAL scenarios the described effects reach such a degree that persons cannot be detected any more when applying the commonly used full body models. Examples of such scenarios are illustrated in Fig. 1.

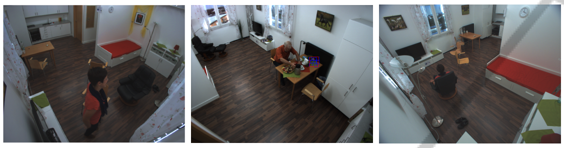

Figure 1: Person detection when using the full body model: Persons standing very close or almost below the camera, occluded by objects such as a table, or with the back turned to the camera cannot be reliably detected. There often occur false positives or false negatives in such cases.

Nevertheless it can be observed that in all those images the head-shoulder part is always visible. For this reason, we propose to use this very part to build a new model for the DPM classifier. By using this Latent SVM based approach, a dynamical assignment of part models to overall models is possible. Since we propose to use the DPM approach with considering detection of the head-shoulder of a person, this method is explained in this section.

The model is characterized by a lower-resolution root filter, several spatially flexible, higher-resolution part filters and a spatial model for the locations of every part with respect to the root. The generated model is shown in Fig. 2.

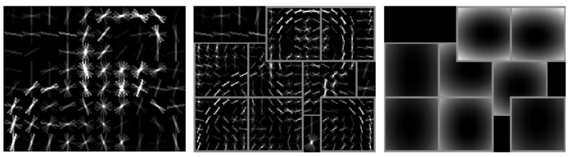

Figure 2: Illustration of a head-shoulder model. The left image shows the lower-resolution root filter, the image in the middle demonstrates higher-resolution part filters and the right image the spatial model.

When a person shall be localized, a feature pyramid is created in the first step by down-scaling the image. After the calculation of HOG features at every scale, the filters are applied in order to get the filter response for every single part model. The part models are then combined at one scale and the results of every scale are finally fused in order to compute a final score for the root locations. Fig. 3 shows the result of the example image (see second image in Fig. 1), where the locations of the found parts and the root are marked by blue and red rectangles respectively.

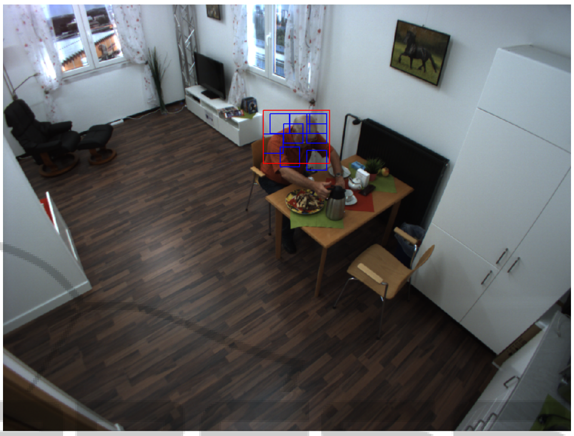

Figure 3: Exemplary detection of the DPM algorithm using the proposed head-shoulder model. When using the head-shoulder model, the person sitting at the table can be detected.

4 DATA PROCESSING CHAIN

In order to get a powerful person model, several steps starting from data acquisition to model evaluation were performed.

4.1 Data Acquisition

For training purposes, a CCD Camera (Allied GC1350) with 1.3 MP and a wide lens were used to cover the whole room. In some scenarios, the camera is mounted at the ceiling with a tilt angle of about 20^o and a height of 2 m from the floor. In other scenarios, the camera is held at a certain height in front of the persons. Generally it is important to build a model in such a way that it shows good performance under different conditions. Therefore the exposure time as well as the gain of the camera were set to different values in order to obtain images recorded under varying lighting conditions.

Altogether, 11920 images (9013 positive and 2907 negative) were collected whereas 13 male and female persons of different age groups participated in the recordings. All persons wore different clothing to get variations in our dataset. In the scenarios, the actors were sitting, standing and lying.

The negative images were randomly collected from the Internet. In these images no persons are present.

4.2 Data Annotation and Database

For data annotation, files in PASCAL style (Everingham et al., 2010) were generated from our data collection. In order to annotate an object in an image, we specified bounding boxes and labelling conditions such as segmented, truncated, occluded, or difficult. Multiple objects from multiple classes may be present in the same image. However, in our dataset, only one object was allowed in the same image.

Therefore, positive images each contain exactly one single person. Every person's head-shoulder part was specified by manually fitting a bounding box to this region in the image. As this process was very time consuming, a labelling tool was developed to increase the annotation speed. It enabled the drawing of the bounding rectangles in the images intuitively.

The annotated data was split into training and testing data sets. For a sensible evaluation, these two datasets were independent from each other.

4.3 Training

A few models were trained to detect both full body and head-shoulder using the DPM and ACF detector.

Both training and testing were performed in Matlab code. The Matlab Toolbox "Piotr's Image and Video Matlab Toolbox" (Dollár, 2013) and an object detection system using deformable part models (DPMs) and latent SVM (voc-release5) (Girshick et al., 2012) (Felzenszwalb et al., 2010) are utilized.

5 EXPERIMENTAL SETUP AND RESULTS

For evaluation, a number of experiments were performed. Firstly, two algorithms were compared while considering the head-shoulder and the full body models. Secondly, we analyzed the performance of the head-shoulder detector in our approach by tuning several parameters. Finally, the performance of the proposed detector was evaluated by a public dataset. Table 1 gives an overview about the experiments.

Precision-recall curves were used to compare the two considered algorithms. False positive rate (FPR) as well as true positive rate (TPR) were determined at different threshold values. If the overlap ratio between the detected and the labelled rectangle was higher than a certain value the sample is counted as true positive else it is a false positive.

Table 1: Experimental overview.

Experiments	Content
Experiment 1	DPM vs. ACF detector full body (our dataset)
Experiment 2	DPM vs. ACF detector head-shoulder (our dataset)
Experiment 3	head-shoulder vs. full body (DPM) human pose (sitting, standing) partial occlusion (our dataset)
Experiment 4	head-shoulder (DPM) distance (near, mid, far) lighting (bright, dark) (our dataset)
Experiment 5	head-shoulder (DPM) number of images used for training (our dataset)
Experiment 6	head-shoulder (DPM) (public dataset)

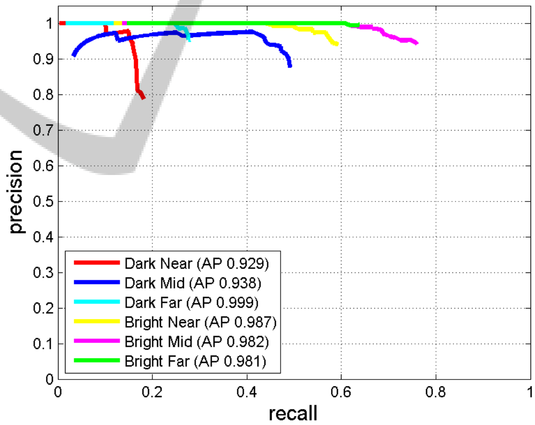

Figure 4: The performance of the full body DPM model (trained using the Pascal dataset, cascade PCA 5, the overlap area > 10%).

5.1 Experiment 1

In the first experiment, we compared the full body DPM and ACF detector deploying a set of test data shown in Table 3.

However, these two detectors have been trained: The DPM was trained with the Pascal person dataset while ACF detector was trained with the Inria dataset. The performances of the full body DPM and ACF detector on the testing data are shown in Fig. 4 and Fig. 5 respectively.

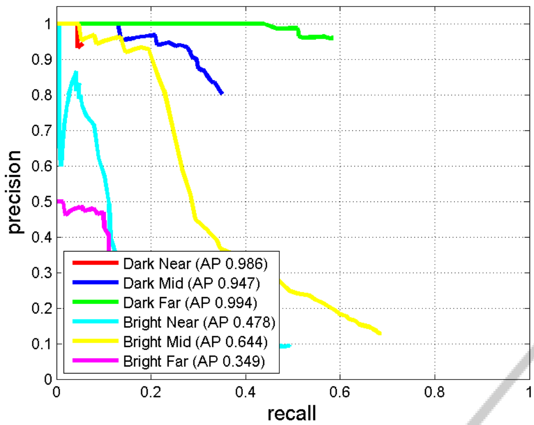

Figure 5: The performance of the full body ACF detector (trained using the Inria dataset, the overlap area > 10%).

5.2 Experiment 2

In this experiment, we trained the models for detecting the head-shoulder part using our own dataset designed in Sect 4. The purpose is to analyze the performance of two different algorithms using the same training and testing data for the head-shoulder. Deploying the testing data shown in Table 3, the performances of the models are shown in Fig. 6 and Fig. 7.

Deploying the ACF detector, different parameters' values are assigned to train several ACF detector models.

5.3 Experiment 3

In this experiment, the full body model was compared against the head-shoulder model. Both have been trained with the DPM approach. The test data used in this experiment consists of 1601 images with three different persons wearing different clothing. They performed actions of sitting (on a chair, armchair and bed), standing or walking. In this scenario, the persons were often partially occluded and not always fully visible. The TPR and FPR results are shown in Table 2.

Table 2: Comparison of the full body and the head-shoulder model.

Model	TPR	FPR
Full body	48 %	1 %
Head-shoulder	93 %	2 %

5.4 Experiment 4

For this experimental setup, we considered two parameters: distance and lighting conditions. There are three distance values (near, mid and far), which describe the distance between a person and the camera. The lighting can be either dark or bright, see Fig. 8.

Table 3: Number of images for the different configurations.

Lighting, Distance	Number of Images
Dark-near	309
Dark-mid	306
Dark-far	338
Bright-near	301
Bright-mid	360
Bright-far	305

Consequently, this results in six different parameter configurations, see Table 3. The number of images used in every configuration is almost equally distributed.

In Fig. 6, the precision and recall curves illustrate the performance results including average precision (AP).

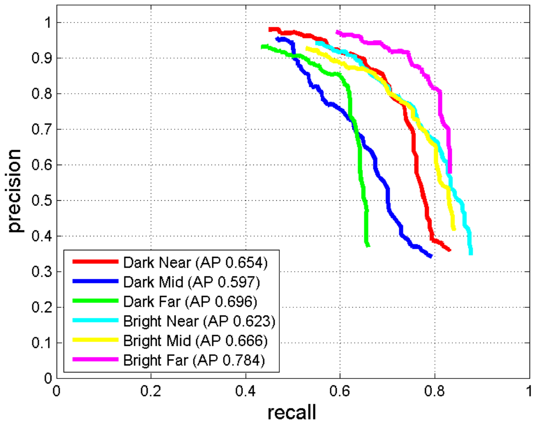

Figure 6: The performance of head-shoulder DPM model (trained using around 9000 positive images, cascade PCA 5, the overlap area > 50%).

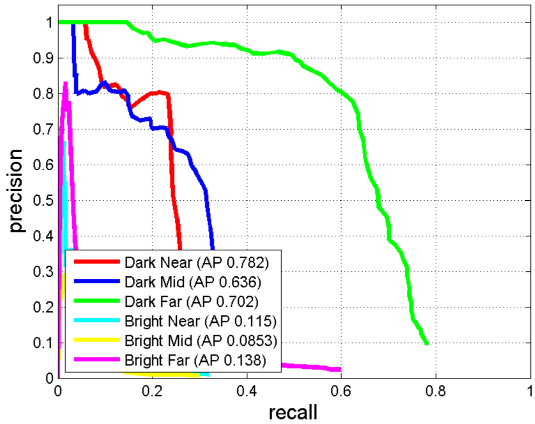

Figure 7: The performance of head-shoulder ACF detector (trained using around 9000 positive images, the overlap area > 50%).

Figure 8: The left and right images are the samples in the test group of dark-mid and bright-near respectively.

5.5 Experiment 5

In this experiment, 9000 positive images were used. Furthermore, another DPM head-shoulder was trained using 1000 positive images for the purpose of comparing the effect of using different number of training data. Fig. fig:exp2-1 and Fig. 9 presents the performance of these models respectively on the test data shown in Table 3.

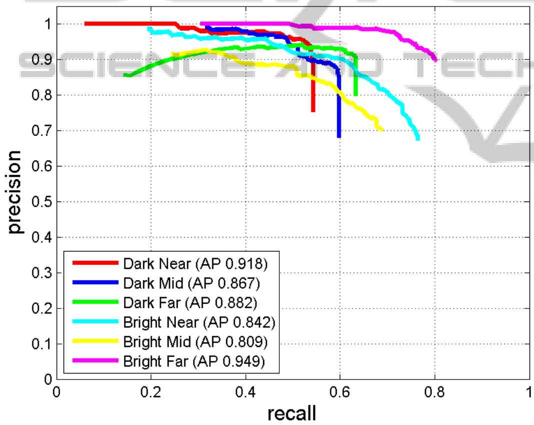

Figure 9: The performance of head-shoulder DPM model (trained using around 1000 positive images, cascade PCA 5, the overlap area > 50%).

5.6 Experiment 6

In the last experiment, we evaluated the performance of head-shoulder DPM model on a public dataset. In that way, we tested whether our trained models are still effective in other AAL scenarios as well. Therefore, the test data set from GER'HOME laboratory (Zouba et al., 2007) was employed. The result is shown in Table 4.

Table 4: Test on public data set.

Images	Model	TPR	FPR
GER'HOME	Full body	76 %	12 %
GER'HOME	Head-shoulder	93 %	2 %

6 DISCUSSION

6.1 DPM vs. ACF Detector

In the first and second experiments (see the comparision of Fig. 4, 5 as well as see Fig. 6, 7), the results show that the DPM models are better than ACF detector for both full body and head-shoulder models.

With regard to the experiments comparing the results of dark far views with bright far views in Fig. 5 and Fig. 7, one can say that the ACF detector fails to detect a person in bright images. It might be due to the different training dataset that was used to train that model. In the ACF detector approach, the distance factor does not play any role for the performance of the model.

6.2 Head-shoulder vs. Full Body

The results in Table 2 show that the TPR of the head-shoulder model is two times higher than the TPR of the full body model. When an appropriate threshold is applied to the detection scores, the DPM head-shoulder works more robustly than the DPM full body model.

The head-shoulder model gives very good results in the following scenarios: parts of the person are occluded while the person is sitting, the person is standing behind other objects and the person is standing very close to the camera. In these cases, the full body model shows poor performance. Furthermore, the full body model often fails to detect the person while standing up, sitting down and bending to pick up an object from on the floor.

Moreover, the head-shoulder model is tested on public datasets and it is proven that the head-shoulder detection has higher detection rates than the full body model (see Table 2 and 4). By means of these experiments it could be demonstrated that the head-shoulder model is more efficient in AAL scenarios than the full body model.

6.3 Distance

In Sect. 5.4, the distance factor in the model performance was evaluated. Distance refers to the persons' positions with respect to the camera. One purpose of the study is to design a person detection model which is independent from the distance.

In Fig. 6 and Fig. 9, the results show that the distance factor has no influence on the head-shoulder model. For instance, even if the person is near, the head-shoulder is still visible and the head-shoulder silhouette is not affected by the perspective distortion.

However, the distance factor has effects on the full body detection. In our observation, the full body model has good precision and higher recall when the person is far, which means that the person's full body is in the view. When the person is near, the full body model has a very low recall due to top view, huge perspective distortion and occlusion of the lower body.

6.4 Lighting

In Fig. 6 and Fig. 9, the precision-recall curve shows that the DPM head-shoulder model performs better under brighter light conditions. In Fig. 4, the recall value reaches 0.8 when the images have a high contrast and the person is far from the camera. However, this value reaches can only 6.2 in the low contrast images. To sum up, we can say that generally, good illumination condition yields a better result without depending on the distance of the person to the camera.

Nonetheless, we believe that the performance relatively depends on the training dataset. If we would have trained a model using more number of low-contrast images, the result might be reversal.

6.5 Number of Training Samples

Two head-shoulder models are trained using a different number of positives. In Sect. 5.5, one part of the experiment is to compare these models. The model with a lower number of positive images (see Fig. 9) has higher precision rates but less recall compared to the model trained with more number of positive images (see Fig. 6). Thus, training a model with more positive images generalizes the parameters. Therefore, more false detection alarms occur when lowering the precision rates (see Fig. 9).

7 CONCLUSIONS

Person detection plays an important role in many applications. For AAL applications, we analyzed the performance of state-of-the art person detection algorithms using full body models. It was proven that they lack robustness especially when parts of the person are occluded, e. g. because the person is standing very close to a tilted camera mounted at the ceiling or if the person is turned with the back to the camera while sitting. We therefore introduced a head-shoulder model, because this part is visible in most cases.

For training and testing purposes, we collected our own data set and annotated it. The performance was analyzed for the DPM and the ACF detector. Furthermore, we compared the efficiency of the head shoulder model and the full body model. The results show that the head-shoulder model is more robust than the full body model in AAL scenarios. Another finding was that the DPM outperforms the ACF detector. We proved on a public dataset that the head-shoulder DPM model is very efficient as well. In addition to that, the detector was successfully tested under different distance and lighting conditions.

In future, the DPM head-shoulder model has to be enhanced with further training samples showing other view points of persons, e. g. lying in the bed. Naturally, this approach could also find usage in a wide range of other application fields, like security, consumer market or public facilities.

ACKNOWLEDGMENTS

This study is funded by the European Fund for Regional Development (EFRE). Special thanks are furthermore expressed to all persons participating in the recording and labelling process.

REFERENCES

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

Dollár, P. (2013). Piotr's Image and Video Matlab Toolbox (PMT). http://vision.ucsd.edu/∼pdollar/toolbox/doc/index.html.

Dollár, P., Belongie, S., and Perona, P. (2010). The fastest pedestrian detector in the west. In *BMVC*.

Dollár, P., Tu, Z., Perona, P., and Belongie, S. (2009a). Integral channel features. In *BMVC*.

Dollár, P., Wojek, C., Schiele, B., and Perona, P. (2009b). Pedestrian detection: A benchmark. In *CVPR*.

Dollár, P., Wojek, C., Schiele, B., and Perona, P. (2012). Pedestrian detection: An evaluation of the state of the art. *PAMI*, 34.

Enzweiler, M. and Gavrila, D. (2009). Monocular pedestrian detection: Survey and experiments. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(12):2179–2195.

Ess, A., Leibe, B., and Van Gool, L. (2007). Depth and appearance for mobile scene analysis. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8.

Everingham, M., Van Gool, L., Williams, C., Winn, J., and Zisserman, A. (2010). Pascal. *International Journal of Computer Vision*, 88(2).

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

Felzenszwalb, P., McAllester, D., and Ramanan, D. (2008). A discriminatively trained, multiscale, deformable part model. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8.

Gavrila, D. (2007). A bayesian, exemplar-based approach to hierarchical shape matching. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(8):1408–1421.

Girshick, R. B., Felzenszwalb, P. F., and McAllester, D. (2012). Discriminatively trained deformable part models, release 5. http://people.cs.uchicago.edu/ rbg/latent-release5/.

Kirchner, N., Alempijevic, A., and Virgona, A. (2012). Head-to-shoulder signature for person recognition. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 1226–1231.

Li, M., Zhang, Z., Huang, K., and Tan, T. (2009). Rapid and robust human detection and tracking based on omega-shape features. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 2545–2548.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110.

Papageorgiou, C. and Poggio, T. (2000). A trainable system for object detection. *International Journal of Computer Vision*, 38(1):15–33.

Park, D., Ramanan, D., and Fowlkes, C. (2010). Multiresolution models for object detection. In *Computer Vision ECCV 2010*, volume 6314 of *Lecture Notes in Computer Science*, pages 241–254.

Richter, J., Findeisen, M., and Hirtz, G. (2014). Assessment and Care System Based on People Detection for Elderly Suffering From Dementia. In *Consumer Electronics Berlin (ICCE-Berlin), 2014. ICCEBerlin 2014. IEEE Fourth International Conference on Consumer Electronics*, pages 59–63. IEEE.

Sabzmeydani, P. and Mori, G. (2007). Detecting pedestrians by learning shapelet features. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–8.

Shashua, A., Gdalyahu, Y., and Hayun, G. (2004). Pedestrian detection for driving assistance systems: single-frame classification and system level performance. In *Intelligent Vehicles Symposium, 2004 IEEE*, pages 1–6.

Tu, J., Zhang, C., and Hao, P. (2013). Robust real-time attention-based head-shoulder detection for video surveillance. In *Image Processing (ICIP), 2013 20th IEEE International Conference on*, pages 3340–3344.

Viola, P. and Jones, M. (2001). Robust real-time object detection. In *International Journal of Computer Vision*.

Viola, P., Jones, M., and Snow, D. (2003). Detecting pedestrians using patterns of motion and appearance. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 734–741 vol.2.

Walk, S., Majer, N., Schindler, K., and Schiele, B. (2010). New features and insights for pedestrian detection. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1030–1037.

Wang, S., Zhang, J., and Miao, Z. (2013). A new edge feature for head-shoulder detection. In *Image Processing (ICIP), 2013 20th IEEE International Conference on*, pages 2822–2826.

Wojek, C. and Schiele, B. (2008). A performance evaluation of single and multi-feature people detection. In *Pattern Recognition*, volume 5096 of *Lecture Notes in Computer Science*, pages 82–91.

Wojek, C., Walk, S., and Schiele, B. (2009). Multi-cue onboard pedestrian detection. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 794–801.

Wu, B. and Nevatia, R. (2005). Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 1, pages 90–97 Vol. 1.

Zhu, Q., Yeh, M.-C., Cheng, K.-T., and Avidan, S. (2006). Fast human detection using a cascade of histograms of oriented gradients. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 1491–1498.

Zouba, N., Bremond, F., Thonnat, M., and Vu, V. T. (2007). Multi-sensors analysis for everyday activity monitoring. *Proc. of SETIT*, pages 25–29.