DETECTING AND TRACKING PEOPLE IN MOTION
A Hybrid Approach Combining 3D Reconstruction and 2D Description

Peter Holzer[1], Chunming Li[2] and Axel Pinz[1]
[1]*Institute of Electrical Measurement and Measurement Signal Processing*
Graz University of Technology, Kronesgasse 5, Graz, Austria
[2]*College of Information Science and Engineering*
Hebei University of Science and Technology, Yuxiang Street 26, Shijiazhuang, China

Keywords: Multibody structure and motion, Fusion of reconstruction and recognition, Person tracking, moving observer.

Abstract: We analyze the most difficult case of visual surveillance, when people in motion are observed by a moving camera. Our solution to this problem is a hybrid system that combines the online 3D reconstruction of stationary background structure, camera trajectory, and moving foreground objects with more established techniques in the 2D domain. Once this 3D part has succeeded in focusing the attention on a particular, moving foreground object, we continue in the 2D image domain using a state-of-the art shape-based person detector, and meanshift-based object tracking. Our results show various benefits of this hybrid approach beyond improved detection rate and reduced false alarms. In particular, each individual algorithmic component can benefit from the results of the other components, by gathering a richer foreground description, improved self-diagnosis capabilities, and by an explicit use of the available 3D information.

1 INTRODUCTION

The past decade has seen many research contributions in high-level vision that have led to a lot of very successful applications in object detection and in surveillance. In particular, person detection is a highly relevant task, with substantial progress and success reported for both, person detection in 2D images, and person tracking by stationary surveillance cameras. In contrast, this paper addresses the much harder problem of tracking people in motion, by an arbitrarily moving observer. In this case, many standard techniques may fail, due to various reasons, including motion blur, permanently changing background conditions, simultaneous background and foreground motion, etc. Other established techniques (e.g. factorization-based approaches to “multibody structure and motion” - MSaM) may be inadequate because they are computationally too expensive to be applied online. We address exactly these issues, analyze strengths and weaknesses of particular algorithms, and propose a novel, hybrid approach that successfully combines online 3D reconstruction by MSaM, reliable 2D person detection by “histogram of oriented gradients” - HOG, and robust 2D tracking by Meanshift. Furthermore, each individual component of our algorithm can benefit from the results of the other components in terms of reduced false positives, improved 3D structure representation, and better self-diagnosis in cases of lost tracking targets.

Related work includes Shape and Motion recovery, object recognition, and 3D structure recovery.

There has been a detailed survey on visual surveillance (Hu et al., 2004) and pedestrian detection (Lopez et al., 2010). Both mainly consider static cameras for video recording.

Person detection methods can be classified into probabilistic-based and non-probabilistic algorithms. Probabilistic-based algorithms segment a person according to a previously established model. (Yan and Pollefeys, 2008) build a kinematic chain of an articulated object to segment articulated motion within non-rigid parts. (Song et al., 2000) give a method based on learning an approximate probabilistic model of the joint positions and velocity of different body features. These methods are effective but more complicated for establishing a model. On the contrary, non-probabilistic methods are more simple and adaptive to many kinds of objects, i.e. they are not limited to human models. Among these methods, HOG-based methods (Dalal et al., 2006; Felzenszwalb et al., 2008; Lin and Davis, 2010) are the current state of art in person detection. (Dalal and Triggs, 2005) use HOG to detect stationary people who are upright and

Holzer P., Li C. and Pinz A..
DETECTING AND TRACKING PEOPLE IN MOTION - A Hybrid Approach Combining 3D Reconstruction and 2D Description.
DOI: 10.5220/0003317005610568
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2011), pages 561-568
ISBN: 978-989-8425-47-8

fully or almost fully visible. Based on this idea, (Lin and Davis, 2010) use deformable part models and a latent SVM to improve the performance. (Felzenszwalb et al., 2008) present an idea of matching a hierarchical part template tree to detect humans and estimate their poses. (Dalal et al., 2006) also combined a human shape descriptor with optical flow to detect moving people from a video. This algorithm runs a detection window across the image at all positions and scales, which is time consuming.

Active tracking of people as well as other objects is challenging. Core tasks are (i) the detection and tracking of rigid or sparsely rigid objects by spatial-temporal trajectories, (ii) the reconstruction of the (unknown) scene structure, and (iii) the pose estimation of the moving observer. Multibody Structure and Motion (MSaM) addresses these issues. In MSaM, (Schindler et al., 2008) distinguish between algebraic methods including factorization-based algorithms (e.g. (Costeira and Kanade, 1995; Costeira and Kanade, 1998; Yan and Pollefeys, 2006)), and non-algebraic methods that combine rigid S+M with segmentation. Non factorization-based methods handling multi-view perspective sequences in dynamic scenes are addressed by (Fitzgibbon and Zisserman, 2000; Li et al., 2007; Schindler et al., 2008; Ozden et al., 2010). But most existing MSaM methods are computationally expensive and thus not applicable in real-time. Online MSaM systems, such as (Leibe et al., 2008) and (Ess et al., 2008) are not purely geometry-based and require quite elaborated object detection algorithms. Furthermore, they are restricted to the processing of certain classes of objects only (cars and people).

We use the online MSaM approach of (Holzer and Pinz, 2010). In contrast to (Leibe et al., 2008) and (Ess et al., 2008), it detects and tracks moving rigid and sparsely non-rigid objects in close to real-time. The approach is geometry-based, and its output is in 3D.

2 ENABLING MODULES

This section reviews the basic components of our person detection and tracking system. It involves salient point detection, MSaM, human shape descriptor, and Meanshift tracking. We use the HOG (Dalal and Triggs, 2005) as human shape descriptor. Salient point detection is required within the MSaM algorithm. Meanshift tracking is used to combine the advantages of MSaM (3D information) and HOG (state-of-the-art human shape detector).

2.1 HOG

(Dalal and Triggs, 2005) compute HOG features for human detection. By using linear and Gaussian-kernel SVMs as classifiers, they report an extensive experimental evaluation. HOG shows superior performance in separating the image patches into human and non-human. It is robust against pose and appearance variations of the pedestrians. Various modifications (Lin and Davis, 2010; Felzenszwalb et al., 2008) exist, which improve its performance. Having excellent detection results, HOG generates false positives on person like structures (e.g. billboards showing persons). Additionally, HOG results are 2D (image plane) only.

In order to compute a person descriptor, a training database with positive and negative examples is needed. A HOG descriptor is computed for each training example. These descriptors are used to train the linear SVM. For testing, similar descriptors are established on testing images and are used as input to the trained SVM to verify whether one or more persons occur in the image or not.

We apply the standard HOG implementation by Dalal and Triggs (Dalal and Triggs, 2005). We provide the whole images as input. So we can get also a false positive rate by the HOG.

2.2 Multibody Structure and Motion

Multibody Structure and Motion (MSaM) enables the (i) detection and tracking of moving objects, (ii) observer pose estimation in a global scene, and (iii) scene reconstruction. The major benefit is that all available information is in 3D, i.e. we gain information on depths and object sizes. In order to have access to 3D information, triangulation is required. Basically, the feature points in one image are compared with feature points in another image through epipolar geometry and some descriptor (eg. cross-correlation). This may either happen through a stereo-camera pair or monocular multi-view perspective sequences.

Basically the observer's pose can be estimated by scene reconstruction. Static, non moving feature points (inliers) represent the background structure. By analyzing outliers (typically noise or object motion), it is feasible to detect and track moving objects.

We use (Holzer and Pinz, 2010) to detect and track rigid or sparsely-rigid moving objects. (Holzer and Pinz, 2010) use 3D outlier information to model motion. In contrast to point cloud matching, the used MSaM method establishes a local coordinate system per object.

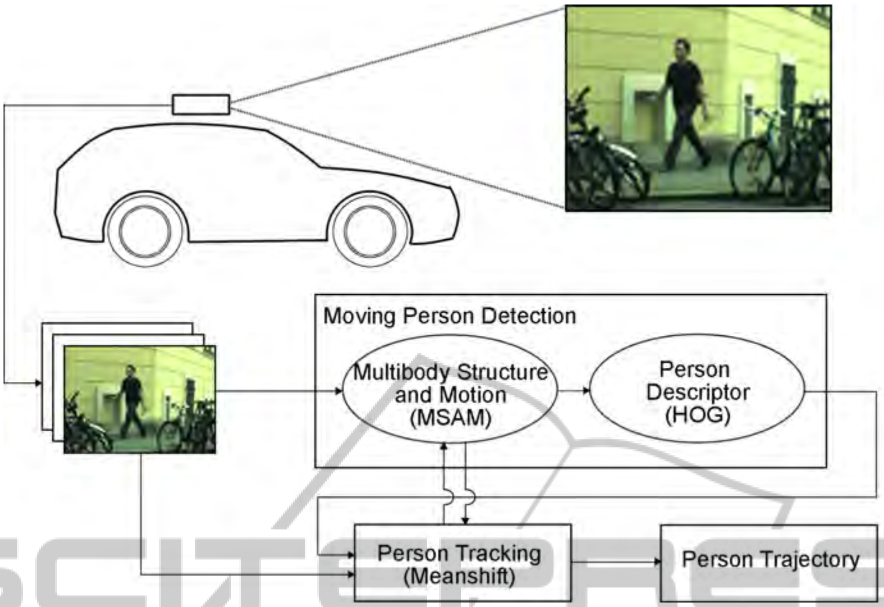

Moving Person Detection
Multibody Structure and Motion (MSAM)
Person Descriptor (HOG)
Person Tracking (Meanshift)
Person Trajectory

Figure 1: Graphical overview of our system. The system can be divided into three main parts: video capture, person detection and person tracking.

2.3 Meanshift Tracking

As both - the person and the observer - are moving, tracking is quite difficult due to the background motion. Meanshift tracking (Comaniciu and Meer, 2002; Comaniciu et al., 2003) is a simple iterative procedure. Its principle bases on a similarity measure. It shifts each data point to the average of data points in its neighborhood. It is efficient for tracking of a large variety of non-rigid objects with different color and/or texture patterns such as human bodies.

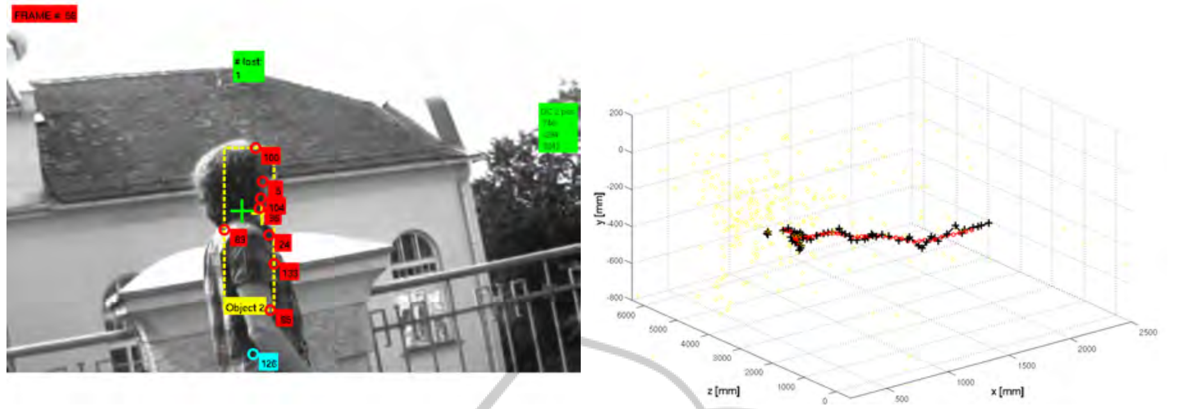
For tracking, Meanshift iterations are used to find the target candidate that is most similar to a given target model. The similarity is expressed by a metric based on the Bhattacharyya coefficient.

3 ROBUST PERSON DETECTION AND TRACKING

In this section, we present our combined detector and tracking method. Our method uses both, motion information and human shape information, to detect and track moving persons. Figure 1 illustrates an overview of our system.

First, MSaM provides us information on moving objects. Then, HOG verifies if the moving object is a person. Finally, Meanshift Tracking is established, to track the moving person. This is a hybrid approach, because Meanshift tracking is established by the combination of HOG and MSaM and the output of these three is compared periodically. Please note, that we do not rely on the results of Meanshift Tracking alone. We rather compare the result of Meanshift (2D) with MSaM (3D projected to 2D image coordinates)/HOG (2D). In case of divergence, i.e. HOG and/or MSaM do not match with the Meanshift tracking any longer, re-initialization of the hybrid tracker is required. Our main contributions are:

- The fast and robust person detector. Multibody moving object detection provides possible locations of persons in 3D. These locations are searched for human shapes. This increases the speed of person detection. Firstly, it can reduce the searching time for a person. The human-shape descriptor (i.e. the HOG) is computed for this subarea only. Secondly, we know the scale because of MSaM. We can limit the scale-pyramid usid in HOG to fewer levels.
- The mutual influence of moving object detection and tracking and person detection makes tracking more reliable. Many false positives detected by the HOG can be eliminated. The output of the hybrid tracking is fed back to the moving object detection (MSaM). There, this information is used to harvest more features on the object. By this, we yield not only outliers but also inliers originally classified as background structure. These additional feature points can be used to further improve the estimation of the moving person's trajectory.



(a) (b)

Figure 2: **(a)** MSaM detection of a moving person. Lost feature points (cyan), active feature points (red), bounding box (yellow), reference point (yellow cross), Kalman prediction (green cross). **(b)** Reconstructed trajectory of the moving object.

3.1 Moving Object Detection

The used MSaM (Holzer and Pinz, 2010) requires feature points in the scene. Stable background features (inliers) are used for scene reconstruction and observer pose estimation. Outliers, which can indicate object motion, are used to detect and track moving objects. We obtain in 3D scene coordinates (i) the reconstructed global scene, (ii) 3D pose information of the observer, and (iii) trajectories of the moving object(s) in 3D. Thus, we know the distance from the observer to the moving object and also the size of the detected (sub-)area.

Estimating the local coordinate system is achieved by analyzing the 3D outlier information. To provide a stable origin, a classification routine separates the available 3D outliers into "active" and "inactive" features. Only "active" features are used for the estimation of the origin. The origin of the local coordinate system is also the reference point of the object; i.e. instead of point cloud matching a single point per object is used to estimate the pose of the object. Additionally, lost features are estimated in case of reappearance. A Kalman filter is introduced, to estimate the position of the object, in case of temporal occlustions.

Our results show that a moving person is detected in most cases. However, in many cases the detected area is smaller than the person (e.g. head only, torso only, etc). Figure 2(a) shows such a basic MSaM detection, figure 2(b) the corresponding trajectory in 3D (top-view).

3.2 Moving Person Validation

The output of the MSaM tracking is validated with HOG. Figure 3 illustrates a correct HOG detection. From MSaM, we know the distance from the observer (camera) to the person. Thus, we know which scale we can apply for the HOG. We cannot guarantee that the output of MSaM covers a complete person, only subparts may be detected instead. But, we can enlarge the MSaM region on the image such that it covers the whole person. The size of the surrounding region can be chosen depending on the distance of the person to the observer. This avoids false positive detections by the HOG.

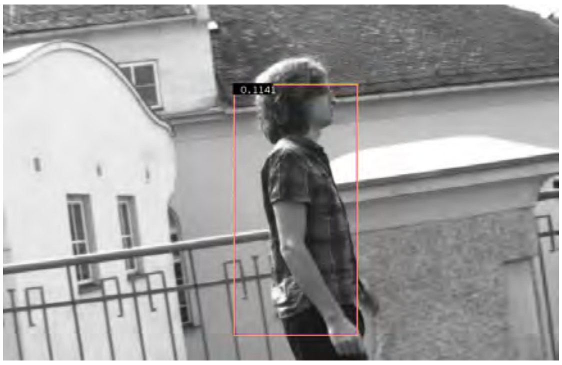


Figure 3: HOG Detection of a moving person.

As the MSaM and HOG detection windows can differ in size massively, we cannot apply the PASCAL criterion here (refer to equation 4). We consider the overlap a_{val} of HOG and MSaM as correct match, if the overlap is larger than 50% of the smaller area of either HOG or MSaM (eg. 1). In most cases, the HOG area is larger, as MSaM mostly detects subparts of a person only.

$$a_{val} := \max\left(a_{MSaM}, a_{HOG}\right) > 0.5 \tag{1}$$

where

$$a_{MSaM} = \frac{area(B_{MSaM} \cap B_{HOG})}{area(B_{MSaM})} \tag{2}$$

$$a_{HOG} = \frac{area(B_{MSaM} \cap B_{HOG})}{area(B_{HOG})} \tag{3}$$

3.3 Supporting Structure by Feedback Control

Once an overlap of HOG and MSaM occurs, Meanshift tracking is initialized. We take the region within the bounding box of the HOG as input for Meanshift Tracking. For the subsequent frames, we consider tracking successful, if either HOG or MSaM overlap with the Meanshift tracking for more than 50%. Otherwise, if for a certain amount of frames neither HOG nor MSaM match with the Meanshift tracking window, Meanshift tracking is stopped. As a human's shape has symmetric properties, it is possible to use Meanshift tracking, i.e. to track a person according to the histogram. In contrast to MSaM, Meanshift Tracking provides 2D information only.

By feeding back the Meanshift tracking information to MSaM, we are in the position to periodically inspect MSaM and Meanshift Tracking. In case of major differences, person tracking is re-initialized.

This feedback routine has also advantages on the available feature points. If MSaM overlaps with the Meanshift tracker, we can search for supporting structure in the overlap. We call every stable feature point (inlier) a supporting structure, if it is in the overlap of MSaM and Meanshift tracker and approximately at the same 3D depth as the object's reference point of the MSaM. With this routine, we gather more feature points on the object, i.e. estimation of the person's trajectory will become more precise. Fig. 4 shows the MSaM tracker, the Meanshift tracker and the gathered supporting structure.

4 EXPERIMENTAL RESULTS

In this section, we present four selected experiments with our hybrid tracking system. These experiments span a range of challenges. Experiment 1 shows a controlled experiment, where a cup and a toy-cow are pulled by a string through the scene. The background contains a lot of structure; amongst other pictures of people. The result demonstrates that our hybrid algorithm can supress false HOG positives. Experiment 2 tracks a person with a fast moving observer. Moreover, the person is not fully in the image. Here, the result of hybrid tracking improves the performance over individual HOG and MSaM. Experiment 3 shows a similar scene, but the person is moving towards the camera, which results in a change of scale. At the end of the sequence, the person is only partially visible. Again the good performance of the hybrid tracking is shown. Experiment 4 is a special case; the person is far away and partly visible behind a set of bicycles. Here, HOG performs much better than MSaM. Neglecting the PASCAL criterion for the hybrid tracking approach, the results are still promising.

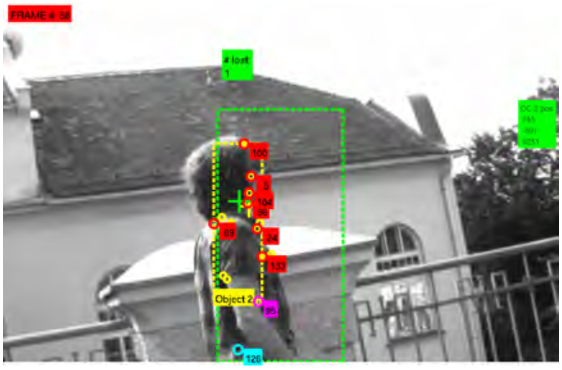


Figure 4: MSaM detection of a moving person (yellow bounding box); Meanshift tracking of person (green bounding box); supporting structure (yellow circles).

When referring to positive detections we consider the PASCAL criterion. This means, (i) the correct detection requires an overlap a_o of the ground truth bounding box B_{gt} and predicted bounding box B_p over 50% and (ii) multiple detections of the same object are considered false detections.

$$a_o := \frac{area(B_p \cap B_{gt})}{area(B_p \cup B_{gt})} > 0.5 \tag{4}$$

The MSaM detections are not evaluated with the PASCAL criterion. As mentioned earlier, most of the detections contain only subparts of an object, depending on the available outlier feature points. We render an MSaM detection correctly, when an object fills at least 50% of the the detected region (equation 1). A correct MSaM detection is illustrated in figure 2(a).

Experiment 1: The scene consists of 180 frames in total. MSaM tracks the moving objects (cup and cow pulled by a string) very well. HOG has no correct detections, as no person is moving in the scene. However, HOG detects 105 false positives in the background. The hybrid approach eliminates the false positives. The hybrid tracking has no results, as no per-

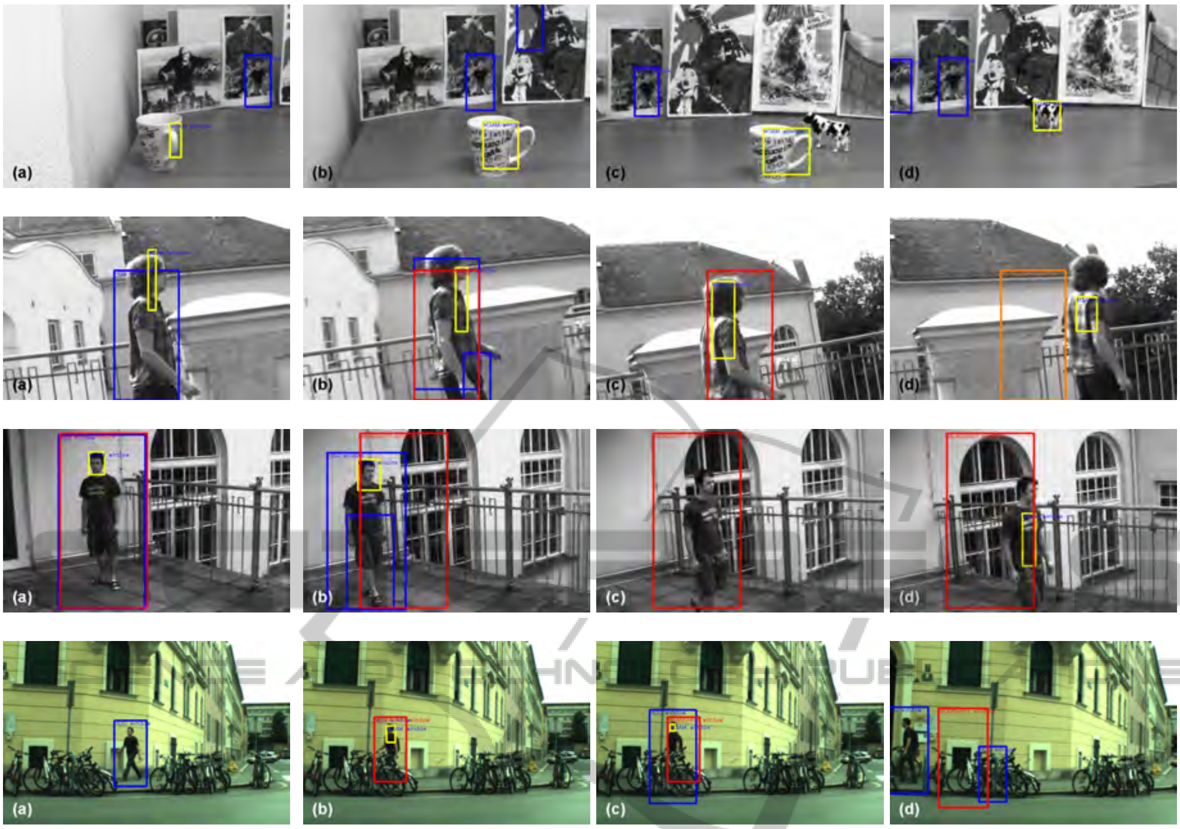


Figure 5: **Experiment 1 (row 1):** 3D-output back-projected to the image-plane. Bounding box of MSaM tracking (yellow); boundig boxes of HOG detections (blue); no hybrid tracking, as no moving person in scene. **Experiment 2 (row 2):** 3D-output back-projected to the image-plane. Overlap of HOG and MSaM initializes hybrid tracking (red) **(a)**; hybrid tracking (red), MSam tracking (yellow), and HOG detections (blue) **(b)**; no HOG detection **(c)**; hybrid tracking lost target (orange), deactivation of hybrid tracking is imminent **(d)**. **Experiment 3 (row 3):** MSaM (yellow), HOG (blue), and hybrid tracking (red) **(a)**; HOG false positive detection, correct MSaM detection (yellow), and false hybrid detections (red) according to the PASCAL criterion **(b)**; only hybrid tracking (red) works **(c)**, correct MSaM tracking (yellow) but false positive detection of hybrid tracking according to the PASCAL criterion **(d)**. **Experiment 4 (row 4):** Only HOG detection (blue) **(a)**; HOG (blue), MSaM (yellow), and hybrid detection (red) **(b)**; HOG (blue), MSaM (yellow), and hybrid tracking (red) works **(c)**, no further MSaM tracking possible, multiple HOG detections (blue), false hybrid tracking (red) **(d)**.

sons move in the scene. Table 1 shows the results. HOG has no positive detections, as no real person is in the scene. Instead, it has a lot of false positives. MSaM tracks the cup reliably (91.5%). The inferior result for the cow (54.5%) is due to the temporal occlusion of the cow by the cup. The hybrid tracking eliminates the false positives of the HOG. As no person is moving, it has no detections. In line “Avg #M gain”, the average amount of additional supporting features gathered by hybrid tracking is listed. As no person is moving, it is equal to zero.

Table 1: Experiment 1: Quantitative Results. 180 frames in total.

	HOG	MSaM	Hybrid
Det. Rate	-	91.5%/54.5%	-
False Pos.	105	-/-	-
No Det.	-	8.5%/45.5%	-
Avg #M gain	-	-	-

Experiment 2: The scene consists of 99 frames in total. The results are shown in table 2. MSaM tracks the person well. The HOG detection rate is rather low, as (i) the observer moves rapidly and (ii) the person is only partly in the scene. MSaM tracking is more reliable, but is below 70% due to motion blur and the lack of outliers on the person in the first 30% of the frames. Hybrid tracking seems to be worse than the MSaM tracking. This is due to the PASCAL criterion. The reqirements on the hybrid tracking are much higher compared to MSaM. Combining the false positives and the correct detecions, hybrid tracking would perform the same as MSaM. 14.1% of no detections

are due to the Meanshift's limits on grayscale images and the too large HOG window on the initialization (a lot of background). With hybrid tracking, we get an average of 8.8 points per frame of additional feature points.

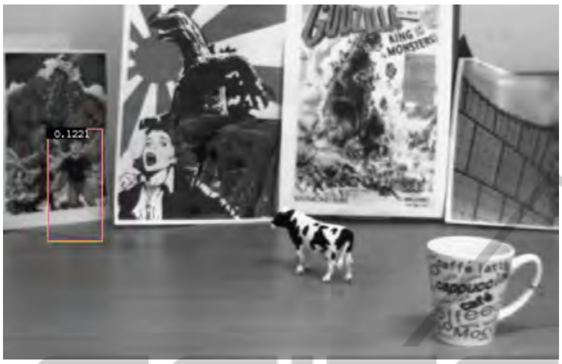


Figure 6: Experiment 1: HOG false positive.

Table 2: Experiment 2: Quantitative Results. 99 frames.

	HOG	MSaM	Hybrid
Det. Rate	17.2%	64.6%	42.4%
No Det.	80.8%	33.4%	14.1%
False Pos.	2	0	21
Avg #M gain	-	-	8.8

Experiment 3: The scene consists of 161 frames in total. In contrast to experiment 2, where the person moves parallel to the observer, here the person walks towards the observer. This results in a scale change of the person. The results are shown in table 3. MSaM tracks the person well. The HOG detection rate is again rather low. MSaM tracking is more reliable, as it does not refer to the PASCAL criterion. Hybrid tracking again seems to be worse than the MSaM tracking. But combining the positive and false detection rate it would outperform the MSaM approach. Feeding back the hybrid tracking result to the MSaM, we get an average amount of 4.6 supporting structure points on the object.

Table 3: Experiment 3: Quantitative Results. 161 frames.

	HOG	MSaM	Hybrid
Det. Rate	23%	79.5%	40.4%
No Det.	77%	20.5%	14.9%
False Pos.	47	-	71
Avg #M gain	-	-	4.6

Experiment 4: The scene consists of 55 frames in total. The results are shown in table 4. The MSaM result is poor. The person is small and uniformly colored, i.e. very few outlier feature points are found on the object. The HOG detection rate is very good, even when the person is partly occluded. Hybrid tracking seems to be worse than the MSaM tracking. Again, neglecting the PASCAL criterion, the result of hybrid tracking is similar to the good performance of the HOG. But in contrast to the HOG, hybrid tracking deals with 3D information. The average amount of 1.1 supporting structure points on the object can be explained be the low hybrid detection rate.

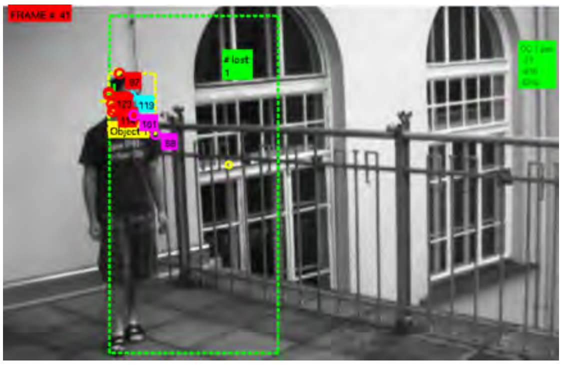


Figure 7: Experiment 1: Meanshift false positive according to the PASCAL criterion.

Table 4: Experiment 4: Quantitative Results. Human in scene: 55 frames.

	HOG	MSaM	Hybrid
Det. Rate	78.2%	16.4%	10.9%
False Pos.	30	0	52
No Det.	20%	83.6%	x
Avg #M gain	-	-	1.1

Summing up all experiments, the following observations can be made:

- The MSaM's detection rate is typically higher than HOG's or hybrid's. As we cannot control, which parts of an object are detected by MSaM (texture), we cannot use the PASCAL criterion.
- The hybrid tracking provides 3D information. We can speed-up the HOG, as (i) we know the distance to the person (fewer pyramid levels) and (ii) we get a rough idea, where to search in an image (region of interest)
- The hybrid tracking provides important feedback for MSaM. We can investigate inliers in a larger subarea (HOG window / Hybrid tracking window). Knowing the distance, we find supporting structure for a person, which can help to improve the estimation of the person's reference point.

5 CONCLUSIONS

We have presented a moving person detection and tracking system. As tracking by a moving observer is a difficult task, we combined 3D algorithms with 2D descriptors and tracking algorithms. The system allows a moving observer and moving objects. Because we use MSaM, we obtain 3D information on the scene, observer motion, and object motion.

By combining different components, we gain a mutual benefit. By combining the HOG with the MSaM tracker, we get 3D information of the person motion and eliminate false postive HOG detections. By feeding back the Meanshift tracking, we can harvest additional features on the object for improved MSaM performance. Our system deals with 3D and 2D information. As we know the 3D depth and the position in the image-plane, we can speed up HOG (fewer pyramid levels, image subarea validation).

Extensions to other categories are possible. The system is not limited to a human shape descriptor. Introducing different descriptors, the system can track different (or even multiple) categories.

REFERENCES

Comaniciu, D. and Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *PAMI*, 24:603–619.

Comaniciu, D., Ramesh, V., and Meer, P. (2003). Kernel-based object tracking. *PAMI*, 25:564–577.

Costeira, J. and Kanade, T. (1995). A multi-body factorization method for motion analysis. In *ICCV*, pages 1071–1076.

Costeira, J. P. and Kanade, T. (1998). A multibody factorization method for independently moving objects. *IJCV*, 29:159–179.

Dalal, N. and Triggs, B. (2005). Histogram of oriented gradients for human detection. In *CVPR*.

Dalal, N., Triggs, B., and Schmid, C. (2006). Human detection using oriented histograms of flow and appearance. In *ECCV*.

Ess, A., Leibe, B., Schindler, K., and van Gool, L. (2008). A mobile vision system for robust multi-person tracking. In *CVPR*.

Felzenszwalb, P. F., McAllester, D., and Ramanan, D. (2008). A discriminatively trained, mulitscale, deformable part model. In *CVPR*.

Fitzgibbon, A. W. and Zisserman, A. (2000). Multibody structure and motion: 3-d reconstruction of indepenently moving objects. In *ECCV*.

Holzer, P. and Pinz, A. (2010). Mobile surveillance by 3d-outlier analysis. In *ACCV Visual Surveillance Workshop*.

Hu, W., Tan, T., Wang, L., and Maybank, S. (2004). A survey on visual surveillance of object motion and behaviors. *Trans. on Systems, Man, and Cybernetics*, 34:334–352.

Leibe, B., Schindler, K., Cornelis, N., and Gool, L. V. (2008). Coupled object detection and tracking from static cameras and moving vehicles. *PAMI*, 30:1683–1698.

Li, T., Kallem, V., Singaraju, D., and Vidal, R. (2007). Projective factorization of multiple rigid-body motions. In *CVPR*.

Lin, Z. and Davis, L. S. (2010). Shape-based human detection and segmentation via hierarchical part-template matching. In *PAMI*.

Lopez, D. M., Sappa, A. D., and Graf, T. (2010). Survey of pedestrian detection for advanced driver assistance systems. *PAMI*, 32:1239–1258.

Ozden, K., Schindler, K., and Gool, L. V. (2010). Multibody structure-from-motion in practice. *PAMI*, 32:1134–1141.

Schindler, K., Suter, D., and Wang, H. (2008). A model-selection framework for multibody structure-and-motion of image sequences. *IJCV*, 79:159–177.

Song, Y., Feng, X., and Perona, P. (2000). Towards detection of human motion. In *CVPR*.

Yan, J. and Pollefeys, M. (2006). A general framework for motion segmentation: Independent, articulated, rigid, non-rigid, degenerate and non-degenerate. In *ECCV*, pages 94–106.

Yan, J. and Pollefeys, M. (2008). A factorization based approach for articulated nonrigid shape, motion, and kinematic chain recovery from video. *PAMI*, 30:865–887.