Independent motion detection with a rival penalized adaptive particle filter

Stefan Becker, Wolfgang Hübner, and Michael Arens

Fraunhofer Institute for Optronics, System Technologies, and Image Exploitation IOSB
Gutleuthausstr. 1, 76275 Ettlingen, Germany

ABSTRACT

Aggregation of pixel based motion detection into regions of interest, which include views of single moving objects in a scene is an essential pre-processing step in many vision systems. Motion events of this type provide significant information about the object type or build the basis for action recognition. Further, motion is an essential saliency measure, which is able to effectively support high level image analysis. When applied to static cameras, background subtraction methods achieve good results. On the other hand, motion aggregation on freely moving cameras is still a widely unsolved problem. The image flow, measured on a freely moving camera is the result from two major motion types. First the ego-motion of the camera and second object motion, that is independent from the camera motion. When capturing a scene with a camera these two motion types are adverse blended together.

In this paper, we propose an approach to detect multiple moving objects from a mobile monocular camera system in an outdoor environment. The overall processing pipeline consists of a fast ego-motion compensation algorithm in the preprocessing stage. Real-time performance is achieved by using a sparse optical flow algorithm as an initial processing stage and a densely applied probabilistic filter in the post-processing stage. Thereby, we follow the idea proposed by Jung and Sukhatme.[1] Normalized intensity differences originating from a sequence of ego-motion compensated difference images represent the probability of moving objects. Noise and registration artefacts are filtered out, using a Bayesian formulation. The resulting a posteriori distribution is located on image regions, showing strong amplitudes in the difference image which are in accordance with the motion prediction. In order to effectively estimate the a posteriori distribution, a particle filter is used.

In addition to the fast ego-motion compensation, the main contribution of this paper is the design of the probabilistic filter for real-time detection and tracking of independently moving objects. The proposed approach introduces a competition scheme between particles in order to ensure an improved multi-modality. Further, the filter design helps to generate a particle distribution which is homogenous even in the presence of multiple targets showing non-rigid motion patterns. The effectiveness of the method is shown on exemplary outdoor sequences.

Keywords: Independent Motion Detection, Particle Filter

1. INTRODUCTION

The detection of moving regions in the field of view of a camera is a central pre-processing step in many vision based systems. In the case of a static camera, a pixel based motion detection can exemplary be determined by frame differencing or other background subtraction methods. A good overview of different background subtraction techniques can be found in[2] or.[3] Another way of extracting motion from a sequence of images is to estimate the optical flow. Motion can be segmented using the characteristics of optical flow vectors in an image sequence, but this method is sensitive to noise and computationally more demanding than the background subtraction method (see[4]).

When capturing the scene with a mobile observer, motion aggregation is more challenging. In such a case, the image flow is not only the result of object motion, but also from the ego-motion of the camera. Moreover these

Further author information:
Stefan Becker: E-mail: stefan.becker@iosb.fraunhofer.de
Wolfgang Hübner: E-mail: wolfgang.huebner@iosb.fraunhofer.de
Michael Arens: E-mail: michael.arens@iosb.fraunhofer.de

Unmanned/Unattended Sensors and Sensor Networks X, edited by Edward M. Carapezza, Panos G. Datskos, Christos Tsamis, Proc. of SPIE Vol. 9248, 92480J · © 2014 SPIE
CCC code: 0277-786X/14/$18 · doi: 10.1117/12.2066635

two major motion types are adverse blended together. In order to detect moving objects robustly, these two independent motions have to be decomposed. Most existing approaches for independent motion detection share the underlying idea of first estimate the flow induced by the mobile observer and then detect parts of image which are not in accordance with the estimated ego-motion.

Existing approaches for estimating the motion of the camera can mainly be divided into three categories: feature-based methods, appearance-based methods, and hybrid methods. Feature-based methods are based on salient and repeatable features that are tracked over subsequent frames. Compared to this sparse sampling strategy, there are appearance-based methods with dense sampling, which use the intensity information of all the pixels in the image or subregions of it. Finally hybrid methods that use a combination of the previous methods. The process of estimating the ego-motion complies with estimating the transformation between two image coordinate system and this is also the problem statement of visual odometry. For a brief overview into different visual odometry methods see.[5] Beside the fact that robust estimation of the ego-motion is an active research area with open challenges, most approaches typically do not estimate the motion of moving objects and do not consider non-rigid motion patterns. These approaches focus on a frame to frame coordinate system transformation, whereas a part of the independent motion detection approaches at least builds the basis for detecting the camera motion with a global parametric transformation for aligning the static background or to systematically examine deviations from the global transformation. Nevertheless visual odometry share some concepts or are the basis of independent motion detection: the more accurate the camera motion is estimated the better the separation of the blended motions. Often additional depth information is used in order to improve the accuracy of the ego-motion estimation and hence the independent motion detection.

Accordingly alternative approaches for independent motion detection differ considerably with respect to the utilized information and their computationally effectiveness. However, the proposed approach is designed to work with an uncalibrated monocular sensor and achieves real-time performance. Like the approach from Ciliberto et al,[6] which also works with a monocular setup. They apply a sparse flow on a uniform dense grid in order to estimate a global motion model. The geometric constraint of the motion model and a neighborhood constraint are use to perform a failure analysis for all grid points. Outliers of their so called cover-uncover trick are interpreted as independent motion. For estimating the outlier of a salient feature based spare flow, Czuni et al.[7] in extension of Fejes et al.[8] used a method based on the projections of the optical flow for monocular observers. By applying several projections of the optical flow and using a voting mechanism to estimate flow vectors which are not consistent with the focus of expansion. Similar to the presented approach, Szolgay et al.[9] use a sparse flow to estimate a parametric global motion model, which is densely applied to calculate a compensated difference image. Instead of a probabilistic filter they use kernel density estimation for further analysis, i.e. for reducing false alarms. Some approaches try to maintain the integrity of a background model even when a sensor moves. The approach of Yi et al.[10] models the background through dual-mode single Gaussian model (SGM) with an age term. The camera motion is compensated by mixing neighboring models when applying the estimated image transformation. Sheigh et al.[11] use the trajectories of salient features to build a sparse representation of the background. Under the assumption of orthographic projection, the 2D trajectories lie in subspace spanned by basis trajectories. The background is subtracted by removing trajectories that lie within the space spanned by the basis. All these approaches use monocular sensors and an ego-motion estimation based on salient features.

The appraoches of Lenz et al.[12] and Gilberto et al.[13] also use a sparse flow but with a stereo camera system. Moreover, Ciliberto et al. use prior knowledge about the kinematic of their humanoid robot to infer the transformation of the sensor platform. In combination with a projection re-projection step using the depth information the ego-motion is predicted and modeled as a Gaussian process. Deviations between tracked features and the propagated features are detected as independent motion. Lenz et al. build a graph-like structure connecting all salient feature tracks with a sparse flow using Delaunay triangulation. The resulting edges are removed according to scene flow differences exceeding a certain threshold with respect to the uncertainty of the computed 3D position of every salient point. In the work of Zhou et al.[14] the salient feature points are reconstructed in 3D via triangulation and reprojected to the image. A motion likelihood for each pixel is obtained by propagating the uncertainty of the ego-motion to the "Residual Image Motion Flow" (RIMF). In order to segment moving objects, the motion likelihood and the depth gradient are used in a graph-cut approach as region and boundary edge weights. The problem of independent motion detection can also be solved by trying to reconstruct the complete scene using structure-from-motion (SfM). Most SfM methods fail in the presence of non-rigid motion

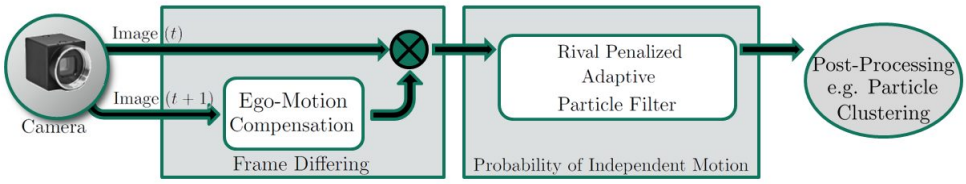

Figure 1. Adaptation of the processing pipeline as proposed by Jung and Sukhatme[1] for detecting multiple moving objects from a mobile monocular camera system with the main moduls ego-motion compensation and the probalistic filter.

patterns or independent motion. There are approaches which adapt SfM for dealing with these problems. They can be categorized as non-rigid multi-body SfM see e.g. Sabzevari et al.[15] for more details. Nevertheless, this method works also for monocular systems but it implies a calibrated sensor.

In this paper, we present an adaptation of the method proposed in Jung and Sukhatme[1] for detecting multiple moving objects from a mobile monocular camera system in an outdoor environment. The motion detection process is performed in two steps: a fast ego-motion compensation algorithm in the preprocessing stage and probabilistic filter in the post-processing stage. In figure 1 the adaptation of the overall processing pipeline is depicted. The proposed approach extends the method of Jung and Sukhatme by an alternative outlier removal strategy in the preprocessing stage and an modified particle filter design in order to better deal with non-rigid motion patterns and ensure an improved multi-modality. In section 2 the ego-motion compensation is described. The proposed design of the particle filter is introduced in section 3. In section 4 a qualitatively evaluation is given. Section 5 provides a conclusion.

2. EGO-MOTION COMPENSATION

In order to detect moving objects, the image flow induced by the observers motion needs to be eliminated. The process of estimating the ego-motion complies with estimating the transformation between two image coordinate systems, which can be done by tracking salient features over subsequent frames. Intuitively, it is clear that the estimation of the ego-motion has to be based on static objects in the environment. Unfortunately, in the case of external motion, the selected corresponding features are contaminated by outliers. Hence, for the transformation model to be estimated accurately, it is important to take an outliers removal strategy into account.

2.1 Feature Selection and Tracking

According to its computational efficiency, the KLT feature tracking algorithm[16] is used for selecting of a set of corresponding features between subsequent frames. In compliance, we picked FAST corners[17] as salient features over other corner detectors (e.g. Harris[18]) or blob detectors like SIFT.[19] For coping with an inhomogeneous salient feature distribution, the corneness threshold is adapted by keeping the strongest interest points in a cell of a lower resolution uniform-sized grid (feature binning, figure 2).

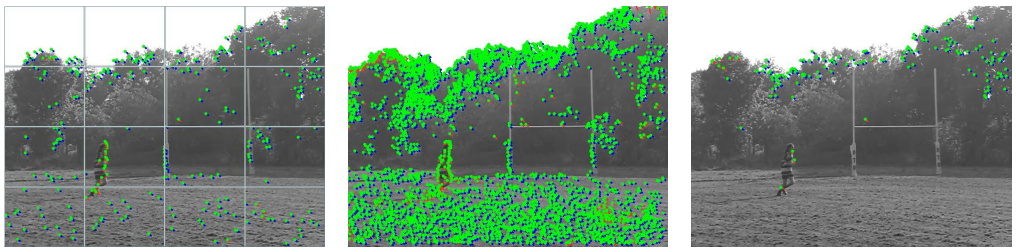

Figure 2. The figures illustrate the strategy of salient feature binning in order to generate a homogeneous distribution. In a lower resolution sized grid the cornerness threshold is adapted by keeping a fixed number of the strongest features in a cell (Left). In the middle, a too low cornerness threshold is chosen and the right figure shows the resulting feature distribution when setting a too strong cornerness threshold.

Under the assumption that the main image flow is associated with the ego-motion, there exists a global parametric transformation which can be used to align the static background. This parametric motion model can be estimated with an optimization method by using two corresponding feature sets F_t and F_{t-1}. Hence, the accuracy of the motion model is a central determining factor for the separation of the involved motion. The complexity of the motion model is a trade-off between a simple, linear model and a highly nonlinear model. When the mobile observer is purely rotating or its translation is relatively slow with respect to the distance between the camera and the background, the change between consecutive images is very small and an affine motion model can cover these transformations. However, when the change increases through a faster observer speed and especially in the case of forward moving observer a nonlinear transformation model is required. For our experiment, we mainly focus on sequences with a relative short interval between two subsequent frames and use an affine global motion model

$$\begin{bmatrix} f_x^t \\ f_y^t \end{bmatrix} = \begin{bmatrix} a_0 f_x^{t-1} & a_2 f_y^{t-1} & a_3 \\ a_4 f_x^{t-1} & a_5 f_y^{t-1} & a_6 \end{bmatrix}.$$

The parameter of the transformation model T_{t-1}^t can be estimated by minimizing the distance $d(F_{t-1}, F_t) = \sum_{i=1}^{N} = (f_i^t - T_{t-1}^t . f_i^{t-1})^2$, where N is the number of corresponding features. As mentioned above, the feature set contains outliers due to external motion. Other common reasons for outliers are image noise, occlusion, blur, strong changes in viewpoint etc.

The strategy of model estimation in the presence of outliers consists of taking advantage of the geometric constrains introduced by the motion model. Robust estimation methods, such as M-estimation,[20] explicitly fitting and removing outliers,[21] and case deletion are established methods in the presence of relative few outliers.[5] Here in order to reduce the effect of features associated with moving objects, we use a two step model estimation procedure with "Random Sample Consensus" (RANSAC).[22] Alternatively or supplementary to the standard RANSAC, a more advanced RANSAC implementation can be applied which estimates the fraction of inliers adaptively.[23] However, we use RANSAC due to its ability to tolerate a tremendous fraction of outliers in order to first compute the initial estimate T_0 using the full feature set F. Then the $\times 84$ outlier rejection rule introduced by Tomasi et al.[24] is applied for partitioning the feature set into a valid subset F_{in} and a rejection or outlier subset F_{out}. For the final estimate T_{t-1}^t is re-computed using again RANSAC with F_{in}. According to Tomasi et al., all features which are more than 5.2 "Median Absolute Deviations" (MADs) away from the median of all distances $d_i = (f_i^t$ - $T f_i^{t-1})^2$ are rejected. Figure 3 exemplary shows the partitioned feature sets: F_{out} is depicted with filled circles and F_{in} with empty circles. Note that in this example most outliers are associated with the independent motion of persons. Although, a robust parameter estimation is applied, it is still assumed that the portion of moving objects in the scene is relatively small compared to static background objects. If this assumptions breaks, e.g. in situations where objects move extremely close to camera, the association of inliers and outliers completely fails. However, if the interchanging keeps in a short time interval and therefore leading to a transient error, the involving probabilistic filter is able to cope with this inlier stealing.

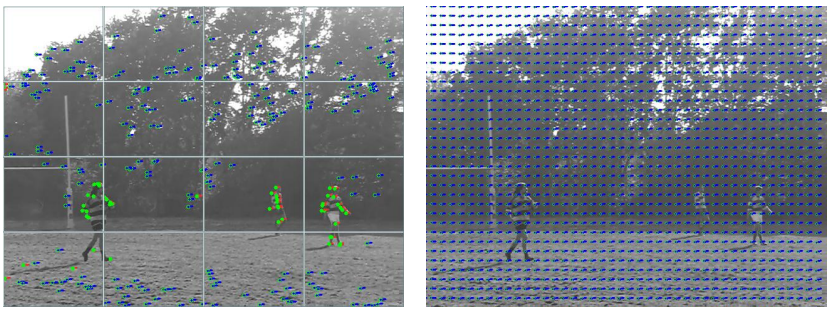

Figure 3. Set of corresponding features. (Left) Outliers rejected with the $\times 84$ rule are marked with filled circles and inliers which are used to estimate the final global transformation model T_{t-1}^t are marked with empty circles. (Right) Densely applied global transformation.

2.2 Frame Differing

In order to generate a motion compensated image I_c, the transformation model T_{t-1}^t is densly applied for each pixel (x, y) of the image I^{t-1}:

$$I_c(x, y) = I^{t-1}(T_{t-1}^t(x, y)) \tag{1}$$

The difference image between two consecutive frames is computed using I_c:

$$I_d = \begin{cases} |I_c(x, y) - I^t(x, y)| \text{ if } (x, y) \in \mathcal{R} \\ 0 \text{ else} \end{cases} \tag{2}$$

Some border pixels are not available in the image I^{t-1}. Hence, the valid region $\mathcal{R}$ of the transformed image is smaller and pixel values outside $\mathcal{R}$ are set to zero. Figure 4 compares the inverted difference images for two consecutive images with ego-motion compensation and without ego-motion compensation. Difference images include higher frequency errors due to the required interpolation for the transformation of the image implicitly involving low-pass filtering. Further errors are due to change in view point or complexity of the scene geometry which could not be predicted with a parametric model. These errors strongly depend on the kind and strength of ego-motion and the scene geometry. Registration errors caused by motion pattern not represented by the motion model are associated with independent moving objects. The results in figure 4 show that in this case despite some errors due to the complex 3D structure of some trees in the background most differences are caused by independent motion patterns.

Figure 4. Examples of difference images from two consecutive frames for the cases of ego-motion compensation on the left and without compensation on the right.

3. INDEPENDENT MOTION DETECTION

The quality of the resulting difference image also depends on the dynamics of the scene and the 3D geometry of the scene. Hence, perfect ego-motion compensation is rarely achievable. However, some of the induced errors can be divided in transient and constant errors over time. In order to filter out transient errors, we follow the idea proposed by Jung and Sukhatme,[1 25] by applying a probabilistic filter in the post-processing stage. The intensity differences originating from a sequence of ego-motion compensated difference images $I_d^0, I_d^1, \ldots, I_d^t$ represent the probability of moving objects. The estimation process is modeled using a Bayesian formulation. Let X_t represent the the state (position and velocity) of a moving object and $P_m(X^t) = P(X^t|I_d^0, I_d^1, \ldots, I_d^t)$ be the posterior probability distribution of the object.

A recursively update of the posterior probability over time by applying a perception model $P(I_d^t|X^t)$ and a motion model $P(X^t|X^{t-1})$, $P_m(X^t)$ can be derived according to the following equation, where η is the normalization constant (see[26] and[25] for more details):

$$
\begin{aligned}
P_m(X^t) &= P(X^t|I_d^0, I_d^1, \ldots, I_d^t) \\
&= \eta_t P(I_d^t|X^t) \int P(X^t|X^{t-1}) P_m(X^{t-1}) \mathrm{d}X^{t-1}
\end{aligned} \tag{3}
$$

For estimating the a posteriori distribution in real time a particle filter,[2728] is used. According to equation 3 a perception model $P(I_d^t|X^t)$ and a motion motion model $P(X^t|X^{t-1})$ are required. The perception model is used to compute the particle weights. Instead of a multi-variate Gaussian the perception model is simplified to a step function with a limited evaluation of a $m \times m$ area. The weight w_i^t of the i_{th} particle $s_i^t = [x_i^t \ y_i^t \ \dot{x}_i^t \ \dot{y}_i^t]^T$ (position and velocity) is computed by

$$
w_i^t = \frac{1}{m^2} \sum_{j=-\frac{m}{2}}^{\frac{m}{2}} \sum_{k=-\frac{m}{2}}^{\frac{m}{2}} I_d^t(x_i^t - j, y_i^t - k). \tag{4}
$$

The Motion model is used to propagate a newly drawn particle according to the estimated motion of a moving object. Without prior knowledge of the object motion, a constant-velocity assumption is made:

$$
s_i^t = \begin{bmatrix} x_i^t + \Delta t \cdot \dot{x}_i^t + \mathcal{N}(0, \gamma_p) \\ y_i^t + \Delta t \cdot \dot{y}_i^t + \mathcal{N}(0, \gamma_p) \\ \dot{x}_i^t + \mathcal{N}(0, \gamma_v) \\ \dot{y}_i^t + \mathcal{N}(0, \gamma_v) \end{bmatrix}. \tag{5}
$$

Here Δt is a time interval and $\mathcal{N}$ is a Gaussian noise term for modeling the uncertainty in position γ_p and velocity γ_v respectively. The additive noise is important to overcome an intrinsic limitation of the particle filter and prevent that all particles move in a convergence direction.
An advantage of using a particle filter is its ability to capture multiple modes of the underlying distribution. Under the following conditions, this ability can ensure a multi target tracking with a single set of particles. On the one hand all objects should be present at the beginning of the estimation process and on the other hand particles should not converge too early onto a single target. Both conditions are not met in most scenarios. Not only the number of independent moving objects is unknown, further most particles are absorbed by the dominant motion and an equivalent amount of motion over time for all objects is not realistic. Especially when dealing with objects showing non-rigid motion patterns, such as the motion of a person, motion of a single object part can dominate the particle filters convergence. As a consequence, the particles are not homogeneously distributed over the objects surface. Such non-rigid motion patterns also impede an acceptable fully segmented object and often leads to a fragmentation into object parts. In a static case this problem can be overcome by using a more sophisticated background subtraction algorithm (see[2] or[3]). Despite the computational cost of a segmentation and the problems of adapting background subtraction algorithm for a mobile observer, we consciously shift this problem to an adapted particle filter design. When relying on segmented blobs there are several approaches to overcome the weakness of consistently maintaining the multi-modality of the target distributions.[29] Moreover under such improved conditions, there exist several sample based designs for a "Single Tracker Multiple Targets" (STMT) problem statement. For example, the approach of Frank et al.[30] in accordance with the "Joint Probabilistic Data Association Filter" (JPDAF) introduced by Bar-Shalom[31] and Formann et al.,[32] which computes the probabilities of measurement association to multiple target tracking. Furthermore, the "Probability Hypothesis Density" (PHD) filter which recursively estimates the number and the state of a set of targets given a set of observations treated as a set-valued state and the collection of observations as a set-valued observation (see[33] or[34]). Unfortunately, the prerequisites for directly applying these concepts are not fulfilled when the probability of moving object does only depend on the difference image. Another way for achieving multi-modality is the concept of "Multiple Trackers Multiple Targets" (MTMT), in order to combine multiple measurements and solve the association problem at the same time (see Blackmann[35]). Multiple trackers are used to track multiple targets cooperatively, but such approaches are also not directly applicable when dealing with non-rigid motion patterns and a division of the object in several parts.

3.1 Rival Penalized Adaptive Particle Filter

The goal of the proposed particle filter design is to generate a particle distribution which is homogenous even in the presence of multiple targets showing non-rigid motion pattern. The overall goal is to ensure an improved multi-modality and to rapidly detect objects entering a scene. As mentioned, in the case of objects showing non-rigid motion patterns, motion of a single object part can lead to very fast convergence on the object part due to their stronger amplitudes in the difference image. The motion of a person is an excellent example, where the limbs of the human body lead to much stronger motion signal. In order to to avoid early convergence on single object parts like arms or legs, the motion signal close to the maximum particle distribution is suppressed by manipulating the difference values. This operation is local, compared to just manipulating the noise parameters γ_p and γ_v which leads to an overall expansion of the particle distribution, not the complete particle set is influenced and hence most of the particle set is unaffected. When $p_{x,y}^{max}$ corresponds to the maximum of the particle distribution at time $t-1$, the pixel values are manipulated according to

$$I_m = \left| I_d^t - I_d^t(x,y) \exp\left(-\frac{(x - T_{t-1}^t p_x^{max})^2 + (y - T_{t-1}^t p_y^{max})^2}{2\gamma_s^2} \right) \right|, \tag{6}$$

where γ_s controls the range of the effective neighborhood. This penalization increases the probability of resampling particles located on the object contour, without complete prevention of a convergence on a uniform moving objects. In order to estimate the maximum of the particle distribution the image is partitioned into an uniform grid, with the cells enclosing the weighted particles. Concurrently, this step is used to build up a k-d tree representation for increasing the efficiency of the filter by adapting the size of the sample set on-the-fly. The key idea of KLD-sampling introduced by Fox et al.[36] is to bound the approximation error caused by the sample-based representation of the particle filter. The approximation error is measured with the Kullback-Leibler distance. When the size of the tree is k, the error bound is ϵ and the confidence quantile is $z_{1-\delta}$, a reasonable number n can be estimated as follows

$$n = \frac{1}{2\epsilon}\chi^2_{k-1,\delta-1} \doteq \frac{k-1}{2\epsilon}\left\{ 1 - \frac{2}{9(k-1)} + \sqrt{\frac{2}{9(k-1)}} z_{1-\delta} \right\}^3. \tag{7}$$

Dynamically adapting the number of particles is not only helpful for achieving real-time performance, but further KLD-sampling provides a good measure on how focused the particle density is. KLD-sampling chooses a high number of particles if the state uncertainty is high and a low number if the density is focused. Accordingly, a dominant motion of an object is present in the scene when the number of particles is low. However, in most cases due to different speed, size and shape of an object, different distances to a camera, etc. there will be a dominant motion when multiple objects are present. In order to capture also the modes of the non dominant moving objects, we use a competition scheme between set of particles. In order to prevent rival particle filters from converging on the same object, the difference image is adaptively modified for subsequent processing whenever a particle filter is updated. In the case of a very strong motion only the pixel values covered by the filter are cleared and the rivals have enough "food" left. When the particle distribution of the first particle set is less focused, the particles of rivals are pushed away due to an increased number of covered pixel values and an adaptively increased area of cleared pixels. Before a rival particle set is updated, the difference image is updated according to

$$I_r^t = |I_m^t - \sum_{l=1}^{n} \sum_{j=p_x^l - \lceil \frac{n}{2n_{max}} \gamma_r \rceil}^{p_x^l + \lceil \frac{n}{2n_{max}} \gamma_r \rceil} \sum_{k=p_y^l - \lceil \frac{n}{2n_{max}} \gamma_r \rceil}^{p_y^l + \lceil \frac{n}{2n_{max}} \gamma_r \rceil} I_m^t(p_x^l - j, p_y^l - k)|, \tag{8}$$

where γ_r is a parameter determining the maximum strength of rival penalization. The ratio $\frac{n}{2n_{max}}$ of the adaptive number of particles n and the maximum number of particles n_{max} for one set is in accordance with the level of convergence related to the dominant motion. In the case where no dominant motion is present, the first particle set mainly eliminates the measurement input of the rival sets and hence their particle weight. With

this filter design the first particle set is always focused on the strongest motion signal. In order to deal with an isolated particle distribution and to rapidly detect objects entering a scene, the concept of an augmented filter design is additionally integrated.[28] The main idea is to add random particles to the particle sets. Thrun et al. introduce this concept of particle injection for dealing with the problem of a kidnapped particle set to add an additional level of robustness for robot localization. However, augmenting particles helps when an object leaves the scene too rapidly to detect new motion patterns. Originally the number of injected random samples is determined by comparing the short-term with the long-term likelihood of sensor measurement. Here the percentage of augmented particles is set to a small fixed level γ_a.

Furthermore, in order to capture all modes of all moving targets in the scene the number of rival particle sets is adjusted dynamically based on the introduced concepts. As mentioned the adaptive estimated number of particles per set measures how focused the corresponding particle distribution is. When the difference between the estimated number and maximum number of particles exceeds a certain threshold a rival set can be released. The new rival uses the modified difference image according to equation 8 as input. The particles of all active filters are summed until the overall particle contingent is exhausted. In this hierarchical system, the measurement is passed from top to down. When the higher rated rivals do not focus and hence do not provide any motion signal left for a certain time the rivals which are no longer required are eliminated.

The posterior probability distribution of all moving objects in the scene is estimated by the weighted particles of all active sets. Nevertheless, there is an explicit assignment which filter deals with stronger motion patterns, the actual number of objects in the scene is not determined. Such an additional post-processing can be done by particle clustering, whereby the generated k-d tree can be used for an effective implementation.[1]

4. EXPERIMENTS

The proposed filter design is qualitatively evaluated on exemplary outdoor sequences. The experiments are done on video sequences from different domains and different sensor platforms. The first experiments are done on a sequence taken with a hand-held camera (640×480 pixel), where a rugby game with several players is captured. Some exemplary results are depicted in figure 5 (top row). On the left the input image is shown. The unweighted particle distribution in the middle, where different rival particle sets have different colors. The starting particle set is marked blue, the second purple. On the left the weighted particle distribution is visualized as a heat map. This visualization scheme is also applied for the other results. All the registration errors from the background structures are filtered out by the proposed approach. The particles are distributed relatively homogeneous over the complete body of the person. In this example, the main parts of the captured scene rely on static objects and the distance between the mobile observer and the background allows a robust ego-motion estimation. The closest person to the camera causes the strongest motion pattern and hence the first particle set is concentrated on it. The motions of the other persons are captured with the second set. The figures in the second row (figure 5) show results from video sequences recorded with a long-wave infrared sensor from a moving vehicle. The vehicle speed is relative slow, so that the nonlinear terms of the transformation model induced by a forward moving sensor are not excessively strong. The strongest motion in the scene is related to a group of persons in the middle, which is captures with the first particle set. In this particular scene three rivals are active in order to capture all modes of the independent motion distribution and to filter out the transient errors. The next sequences are taken from publicly available datasets. The third row of Figure 5 depicts results from the zooming Pan-Tilt-Zoom sequence provided by Goyette et al.[37] Due to the lower resolution (320×240 pixel) the particle range is reduced. Here, the global image flow is induced only by the zoom of the camera instead of an ego-motion of the observer. The interval between consecutive frames is small enough, so that the occurring errors from registration and from interpolation have no sustainable effect and particles can focus on the moving person. The bottom row of figure 5 shows results on the Hopkins 155 dataset.[38] In the particular scenario a walking person is capture with a moving mobile camera with a resolution of 640×480 pixels. Here, the first particle set converges on the walking person. The modified difference image includes no strong motion pattern and hence the rival particle set is not focusing.

All the presented video sequences comply well to the preconditions for the introduced processing pipeline. In all cases the static background can be mainly aligned with a parametric global motion model and the translation of the mobile observer is relative slow. However, the post-processing is not able to deal with persisting registration

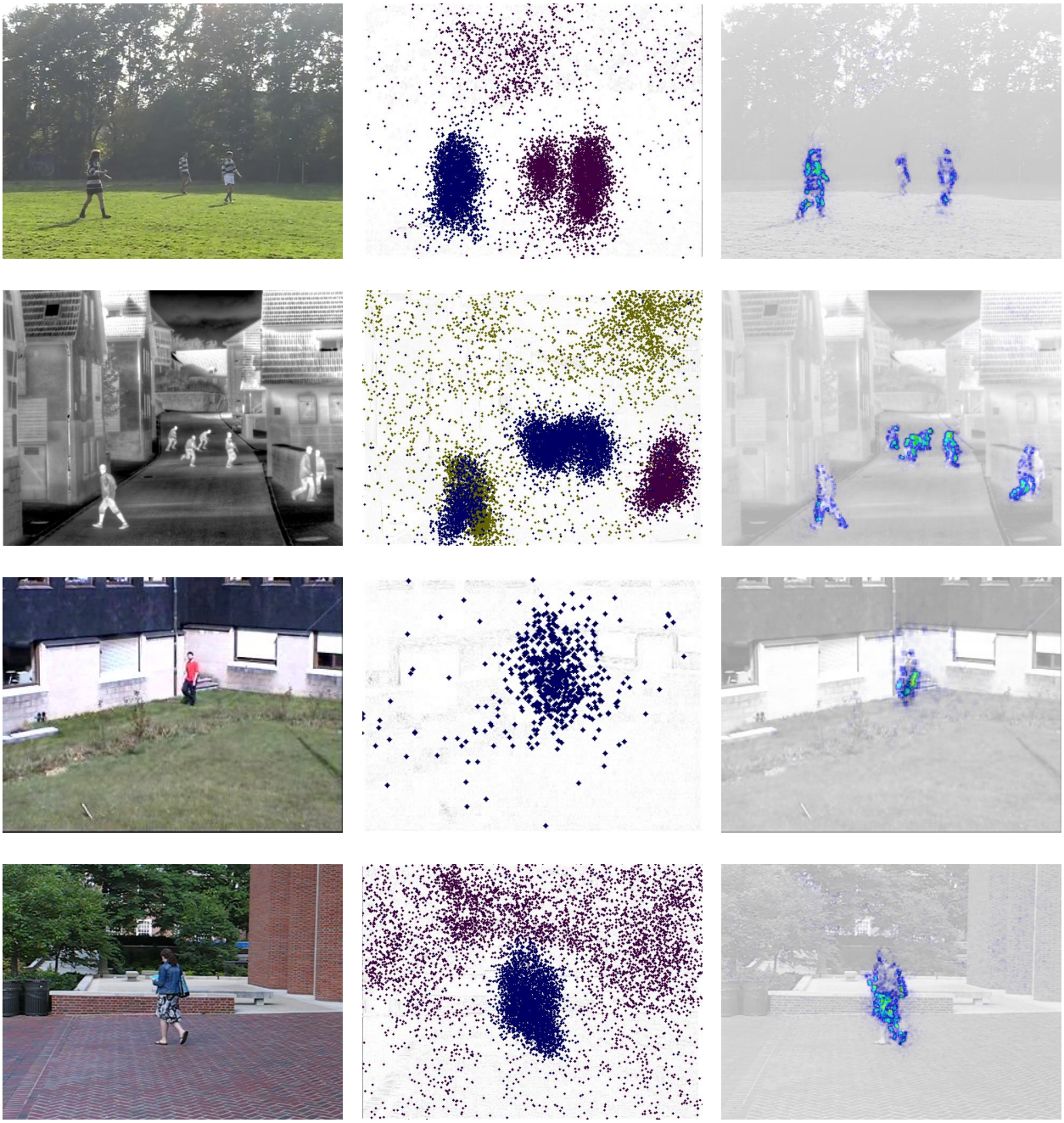

Figure 5. The figures illustrate result from the proposed approach on video sequences from different domains and different mobile observer. From left to right the input image, unweighted particle distribution, and the particle density is shown. The first row shows a video sequence recorded with an handheld camera capturing people playing rugby. Underneath, a video taken with long-waves infrared sensor on a vehicle, where several people cross the street. Then results from the zooming sequence from changedetection dataset provided by Goyette et al.[37] And the figures in last row show results on the sequence "people1" from the Hopkins 155 dataset[38]

errors in the difference image due to strong scene dynamics or a complex 3D geometry of the scene, where a global transformation model is not sufficient.

5. CONCLUSION

In this paper, we presented an adaptation of the method proposed by Jung and Sukhatme[1] for detecting multiple moving objects from a mobile monocular camera system in an outdoor environment. The pipeline consists of a sparse optical ow algorithm for fast ego-motion compensation in the preprocessing stage. In the post-processing stage a densely applied probabilistic filter was used. The proposed design of the probabilistic filter introduces a competition scheme between particles in order to ensure an improved multi-modality. Moreover, the design helps to generate particle distribution which is homogenous even in the presence of multiple targets showing non-rigid motion patterns. The effectiveness of the proposed approach has been shown on video sequences from different domains and captured from different mobile observers. In addition, we discussed several effects, which can lead to non-optimal output and further improvement.

REFERENCES

1. Jung, B. and Sukhatme, G. S., "Real-time motion tracking from a mobile robot," *International Journal Social Robotics* **2**(1), 63–78 (2010).
2. Piccardi, M., "Background subtraction techniques: a review," in [*International Conference on Systems, Man and Cybernetics (SMC)*], 3099–3104 (2004).
3. Sobral, A. and Vacavant, A., "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos," *Computer Vision and Image Understanding* **122**(0), 4 – 21 (2014).
4. Baker, S., Roth, S., Scharstein, D., Black, M., Lewis, J. P., and Szeliski, R., "A database and evaluation methodology for optical flow," in [*International Conference on Computer Vision (ICCV)*], 1–8 (2007).
5. Scaramuzza, D. and Fraundorfer, F., "Visual odometry [tutorial]," *Robotics Automation Magazine* **18**(4), 80–92 (2011).
6. Ciliberto, C., Pattacini, U., Natale, L., Nori, F., and Metta, G., "Reexamining Lucas-Kanade method for real-time independent motion detection: Application to the icub humanoid robot," in [*International Conference on Intelligent Robots and Systems (IROS)*], 4154–4160 (2011).
7. Czúni, L. and Gál, M., "Directional votes of optical flow projections for independent motion detection," in [*International Conference Computer Vision and Graphics (ICCVG)*], 329–336 (2012).
8. Fejes, S. and Davis, L. S., "Detection of independent motion using directional motion estimation," *Computer Vision and Image Understanding* **74**(2), 101 – 120 (1999).
9. Szolgay, D., Benois-Pineau, J., Mgret, R., Gastel, Y., and Dartigues, J.-F., "Detection of moving foreground objects in videos with strong camera motion.," *Pattern Analysis and Applications* **14**(3), 311–328 (2011).
10. Yi, K. M., Yun, K., Kim, S. W., Chang, H. J., Jeong, H., and Choi, J. Y., "Detection of moving objects with non-stationary cameras in 5.8ms: Bringing motion detection to your mobile device," in [*Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*], 27–34 (2013).
11. Sheikh, Y., Javed, O., and Kanade, T., "Background subtraction for freely moving cameras," in [*International Conference on Computer Vision (ICCV)*], 1219–1225 (2009).
12. Lenz, P., Ziegler, J., Geiger, A., and Roser, M., "Sparse scene flow segmentation for moving object detection in urban environments," in [*Intelligent Vehicles Symposium (IV)*], (2011).
13. Ciliberto, C., Fanello, S., Natale, L., and Metta, G., "A heteroscedastic approach to independent motion detection for actuated visual sensors," in [*International Conference on Intelligent Robots and Systems (IROS)*], 3907–3913 (2012).
14. Zhou, D., Fremont, V., Quost, B., and Wang, B., "On modeling ego-motion uncertainty for moving object detection from a mobile platform," in [*Intelligent Vehicles Symposium Proceedings*], 1332–1338 (2014).
15. Sabzevari, R. and Scaramuzza, D., "Monocular simultaneous multi-body motion segmentation and reconstruction from perspective views," in [*IEEE International Conference on Robotics and Automation (ICRA)*], 23–30 (2014).
16. Baker, S. and Matthews, I., "Lucas-kanade 20 years on: A unifying framework," *International Journal of Computer Vision* **56**(3), 221–255 (2004).
17. Rosten, E. and Drummond, T., "Machine learning for high-speed corner detection," in [*European Conference on Computer Vision (ECCV)*], 430–443 (2006).

18. Harris, C. and Stephens, M., "A combined corner and edge detection," in [*Alvey Vision Conference*], 147–151 (1988).
19. Lowe, D. G., "Distinctive image features from scale-invariant keypoints," *International Journal Computer Vision* **60**, 91–110 (2004).
20. Torr, P. and Murray, D., "The development and comparison of robust methods for estimating the fundamental matrix," *International Journal of Computer Vision* **24**(3), 271–300 (1997).
21. Sim, K. and Hartley, R., "Recovering camera motion using l∞ minimization," in [*Conference on Computer Vision and Pattern Recognition (CVPR)*], 1230–1237 (2006).
22. Fischler, M. A. and Bolles, R. C., "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM* **24**(6), 381–395 (1981).
23. Raguram, R., Frahm, J.-M., and Pollefeys, M., "A comparative analysis of ransac techniques leading to adaptive real-timerandom sample consensus," in [*European Conference on Computer Vision (ECCV)*], 500–513, Springer-Verlag, Berlin, Heidelberg (2008).
24. Tommasini, T., Fusiello, A., Trucco, E., and Roberto, V., "Making good features track better.," in [*Conference on Computer Vision and Pattern Recognition (CVPR)*], 178–183 (1998).
25. Jung, B. and Sukhatme, G. S., "Detecting moving objects using a single camera on a mobile robot in an outdoor environment," in [*International Conference on Intelligent Autonomous Systems*], 980–987 (2004).
26. Thrun, S., Fox, D., Burgard, W., and F., D., "Robust monte carlo localization for mobile robots," *Artificial Intelligence* **128**(1-2) (2001).
27. Isard, M. and Blake, A., "Condensationconditional density propagation for visual tracking," *International Journal of Computer Vision* **29**(1), 5–28 (1998).
28. Thrun, S., Burgard, W., and Fox, D., [*Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*], The MIT Press (2005).
29. Vermaak, J., Doucet, A., and Perez, P., "Maintaining multimodality through mixture tracking," in [*International Conference on Computer Vision (ICCV)*], 1110–1116 vol.2 (2003).
30. Frank, O., Nieto, J., Guivant, J., and Scheding, S., "Multiple target tracking using sequential monte carlo methods and statistical data association," in [*RSJ International Conference on Intelligent Robots and Systems (IROS)*], (2003).
31. Bar-Shalom, Y. and Fortmann, T. E., [*Tracking and Data Association*], Academic Press Professional, Inc., USA (1988).
32. Fortmann, T. E., Bar-Shalom, Y., and Scheffe, M., "Sonar tracking of multiple targets using joint probabilistic data association," *Journal of Oceanic Engineering* **8**(3), 173–184 (1983).
33. Pace, M., *Stochastic models and methods for multi-object tracking*, PhD thesis, Université Sciences et Technologies-Bordeaux I (2011).
34. Maggio, E., Piccardo, E., Regazzoni, C., and Cavallaro, A., "Particle phd filtering for multi-target visual tracking," in [*International Conference on Acoustics, Speech and Signal Processing (ICASSP)*], **1**, 1101–1104 (2007).
35. Blackman, S. S. and Popoli, R., [*Design and analysis of modern tracking systems*], Artech House radar library, Artech House, Boston, London (1999).
36. Fox, D., "Kld-sampling: Adaptive particle filters," in [*Advances in Neural Information Processing Systems*], 713–720 (2001).
37. Goyette, N., Jodoin, P., Porikli, F., Konrad, J., and Ishwar, P., "Changedetection.net: A new change detection benchmark dataset," in [*Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*], 1–8 (2012).
38. Tron, R. and Vidal, R., "A benchmark for the comparison of 3-d motion segmentation algorithms," in [*Conference on Computer Vision and Pattern Recognition*], 1–8 (2007).