CZECH TECHNICAL UNIVERSITY IN PRAGUE

Faculty of Electrical Engineering

Diploma Thesis

Jiří Pytela

Identity Preservation for Multi-Target Long Term Tracking

Zachování identity objektů při jejich dlouhodobém sledování ve videu

Department of Cybernetics

Thesis supervisor: **Ing. Matěj Šmíd**

České vysoké učení technické v Praze
Fakulta elektrotechnická

Katedra kybernetiky

ZADÁNÍ DIPLOMOVÉ PRÁCE

Student: Bc. Jiří P y t e l a

Studijní program: Kybernetika a robotika (magisterský)

Obor: Robotika

Název tématu: Zachování identity objektů při jejich dlouhodobém sledování ve videu

Pokyny pro vypracování:

1. Seznamte se s doporučenou literaturou.
2. Rozšiřte stávající systém pro sledování více objektů vyvinutého na CMP o modul pro zachování identity.
3. Připravte experimentální data pro systém sledování více objektů.
4. Otestujte vyvinuté algoritmy, výsledky zhodnoťte, diskutujte selhání navržených algoritmů a kroky k jejich odstranění.

Seznam odborné literatury:

[1] Fleuret, F.; Berclaz, J.; Lengagne, R.; Fua, P.: Multi-Camera People Tracking with a Probability Occupancy Map, PAMI, 2008.
[2] Berclaz, J.; Fleuret, F.; Türetken, E.; Fua, Multiple Object Tracking Using K-Shortest Paths Optimization, PAMI, 2011.
[3] Shitrit, H. Ben; Berclaz, J.; Fleuret, F.; Fua, P.: Multi-Comodity Network Flow for Tracking Multiple People, PAMI, 2013.

Vedoucí diplomové práce: Ing. Matěj Šmíd

Platnost zadání: do konce zimního semestru 2015/2016

L.S.

doc. Dr. Ing. Jan Kybic
vedoucí katedry

prof. Ing. Pavel Ripka, CSc.
děkan

V Praze dne 24. 9. 2014

Czech Technical University in Prague
Faculty of Electrical Engineering

Department of Cybernetics

DIPLOMA THESIS ASSIGNMENT

Student: Bc. Jiří P y t e l a

Study programme: Cybernetics and Robotics

Specialisation: Robotics

Title of Diploma Thesis: Identity Preservation for Multi-Target Long Term Tracking

Guidelines:

1. Study recommended state of the art literature.
2. Design and implement an identity preservation module for the CMP multi-target multi-camera tracking framework.
3. Setup a new experimental data for the CMP tracking framework.
4. Evaluate the implemented algorithms, analyze the results and discuss the failure cases and possible future improvements.

Bibliography/Sources:

[1] Fleuret, F.; Berclaz, J.; Lengagne, R.; Fua, P.: Multi-Camera People Tracking with a Probability Occupancy Map, PAMI, 2008.
[2] Berclaz, J.; Fleuret, F.; Türetken, E.; Fua, Multiple Object Tracking Using K-Shortest Paths Optimization, PAMI, 2011.
[3] Shitrit, H. Ben; Berclaz, J.; Fleuret, F.; Fua, P.: Multi-Comodity Network Flow for Tracking Multiple People, PAMI, 2013.

Diploma Thesis Supervisor: Ing. Matěj Šmíd

Valid until: the end of the winter semester of academic year 2015/2016

L.S.

doc. Dr. Ing. Jan Kybic
Head of Department

prof. Ing. Pavel Ripka, CSc.
Dean

Prague, September 24, 2014

Prohlášení autora práce

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

Dne:

Podpis:

Acknowledgements

I would like to thank Ing. Matěj Šmíd for his supervision and guidance during the work on this thesis. I am grateful that he always found time to discuss the problems I encountered and was as patient with me as he was. I also thank prof. Ing. Jiří Matas Ph.D. for helping me find this interesting topic to work on and for his many suggestions. Most importantly, I would like to express my gratitude to my family, especially my parents, for their support during all the time spent at the university.

Abstrakt

Hlavním cílem této práce je zlepšení přesnosti použitého systému pro sledování objektů ve videu. Vzhledem k tomu, že systém se skládá z více modulů, celkového zlepšení lze dosáhnout úpravami jeho jednotlivých částí.

Prvním krokem při zpracování videa je background subtraction. Zaměřili jsme se na problém falešných detekcí, takzvaných duchů, způsobených nepravidelně se pohybujícími objekty. Vyvinuli jsme metodu pro rozšíření background subtraction algoritmu o detekci duchů, která umožňuje tyto nesprávné detekce identifikovat a odstranit. Navrhli jsme pět různých kriteriálních funkcí, z nich čtyři jsou založené na přítomnosti hran v obraze na hranicích detekovaných objektů a jedna na rozdílu v barevné charakteristice hranice objektu vůči pozadí.

Tuto metodu jsme otestovali jednak vizuálně a jednak pomocí groundtruth pro background subtraction. Rovněž jsme otestovali její vliv na sledovací systém jako celek. Výsledky experimentů prokázaly pozitivní efekt navržené metody jak na samotný výstup background subtraction, tak i na celkový výkon systému.

Analýzou výstupu sledovacího systému jsme určili situace kdy dochází k chybám nejčastěji, to je přítomnost objektů mimo vyhodnocovanou oblast a nebo na krajích obrazu získaného z kamery. K vyřešení těchto problémů jsme provedli úpravy sledovacího systému, konkrétně rozšíření vyhodnocované oblasti a vhodnější pravidlo pro určení viditelnosti objektů v jednotlivých kamerách. Navrhovaná řešení jsme ověřili vyhodnocením jejich vlivu na sledovací systém. Výsledky dokládají významné zlepšení fungování systému.

Všechny experimenty byly provedeny na indoor i outdoor video sekvencích. Problémy, kterými jsme se v této práci zabývali, jsou běžné a nastávají i v mnoha dalších aplikacích. Proto jsme se zaměřili na to, aby i řešení, která jsme vypracovali, byla obecná a tudíž použitelná pro dosažení lepších výsledků v různých aplikacích.

Klíčová slova: Sledování pohyblivých objektů, sledování více kamerami, sledování více objektů, detekce pozadí, detekce duchů

Abstract

This thesis presents work on improving a multi-view multi-target tracking system. The system is composed of separate modules and its overall performance can be improved by modifying the individual modules.

Background subtraction is the first step in processing the video data. The work is focused on solving a problem of false detections, so called ghosts, caused by infrequently moving objects. We developed a general method to identify and avoid these false detections. Different criteria functions for the ghost detection were proposed, four of them based on edge presence on the borders of objects and one based on comparison of the color characteristics on the border of objects to the background.

This method was evaluated visually and using background subtraction groundtruth. The effect of the method on the overall system performance was also tested. In both cases the results show improved performance.

By analyzing the output of the tracking system we identified situations in which errors occur most frequently. Those situations are when objects are present either outside of the area of interest for tracking or on the edge of the field of view of a camera. In order to solve these issues, we proposed modifications of the probabilistic occupancy map module of the tracking system. The modifications consist of extending the area of interest and employing more suitable object visibility rule. The impact of the modifications on the tracking system's performance was evaluated showing significant improvement.

All tracking experiments were conducted on both indoor and outdoor datasets. The problems we faced are common in many applications and the elaborated solutions can be successfully applied to improve them as well.

Keywords: Multi-view tracking, multi-camera tracking, multi-target tracking, multiple object tracking, people tracking, background subtraction, ghost detections

Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation

Object tracking in video is the process of continuously locating an object during a period of time using image data. As a result of lowering costs of computers and availability of inexpensive video cameras, video surveillance has grown significantly from being used by only few companies to protect highly sensitive areas to a widely used crime prevention and security tool.

Surveillance systems are used in most modern cities and the number of surveillance cameras in public places is rising constantly. Manual supervision of the amounts of data these systems provide requires significant effort, it is time consuming and expensive. It is clear that an automated system is necessary in order to efficiently operate complex monitoring systems.

Automated tracking systems have many potential applications beside surveillance systems. Detecting suspicious activity, for example abandoning luggage or aggressive behavior, in controlled areas such as airports or railway and subway stations. Accurately locating people is also important for intelligent environments. In traffic monitoring, such automatic system could be used to help the authorities to identify illegally parked vehicles and other violations. Today, most shopping malls already have video surveillance systems and could analyze the movement patterns of the customers to improve their shopping experience.

In team sports, a tracking system can be a valuable tool for the coach to objectively analyze the players' performance and to optimize the game strategy. It could be also used to improve the viewer experience from the game.

Multi-view object tracking is especially relevant when an extensive area has to be covered, in crowded scenes for resolving object occlusions and for tracking objects in 3D.

We have a working system for multi-view multiple object tracking, but there are many errors occuring during the tracking process. The problems we want to focus on in this thesis are:

- Missing detections
- Double detections
- Identity switches
- Infrequent object movement

1.2 Tracking in Team Sports

Tracking players in team sports has specific issues, different from general people tracking. Recognizing different players based on their appearance is difficult when they are wearing the team uniform. Although, when available, the numbers on the players' uniforms can be used to identify single players. The players frequently change their motion, making their trajectories unpredictable, and close interactions between different players are common. On the other hand, many team sports take place in an environment with controlled light conditions.

1.3 Approach

Our tracking system is composed of three main modules - background subtraction, probabilistic occupancy map computation and k-shortest paths optimization. We believe that the overall performance of the system can be significantly improved by analyzing the results on available data, identifying the problems and improving the single modules to avoid them.

Background subtraction is the first step in processing the video information. When background subtraction was performed on video sequences in order to evaluate the system the results were not satisfactory. By analyzing the problems occuring in our tracking system - missed detections, false detections and identity switches - we were able to identify the situations in which they arise most frequently, and found that often the most efficient way to prevent these mistakes is by improving the background subtraction process.

Moreover, all errors created during the background subtraction step propagate through the entire tracking process, the only way to avoid them is to improve the background subtraction. Achieving the best possible results from background subtraction and handling the problems that often occur during the process is a general problem in computer vision. That is the reason why a large part of this work is dedicated to background subtraction.

2 Data

2.1 Object Tracking Data

To evaluate the tracking system we use two sets of video data:

- **Floorball**
  Video of indoor scene, a floorball match shot from 8 stationary cameras, all with resolution 960x768. Total length of the sequences is 7 minutes, recorded at 20 fps. The groundtruth is available in every 40th frame for the first 1000 frames. The sequence contains the gameplay as well as substitutions. There are 21 unique players and during the game 12 of them are inside the playing field.

  All players wear shorts of different colors and shirts serve as uniforms of their team - black shirts for one team and dark blue shirts for the other team. The result is that different players may have very similar apperance, making correct tracking challenging.

- **PETS'09**
  Video from PETS'09[1][2] dataset containing outdoor scene of people walking in a university campus which serves as simulation of a real-world environment. 7 different cameras are available, 3 of them have a high vantage point and contain the entire scene, 4 of them are positioned at the eye level of a standing person and view only part of the scene. Two cameras record with resolution 768x576, five cameras with resolution 720x576. The sequences are 795 frames long taken at approximatelly 7 fps, groundtruth is available for all frames.

  There is about 10 people present in the scene at any time during the sequence. They wear clothes of different colors.

2.2 Background Subtraction Data

- **Parking**
  Video sequence of a parking car captured on a security camera [3]. There is no groundtruth available for this sequence.

- **Change Detection**
  Videos from 2012 Change Detection[4][5] dataset containing sequences with wide range of detection challenges like dynamic background, camera jitter, intermittent object motion, shadows and thermal signatures. The dataset contains videos of both indoor and outdoor scenes. Groundtruth is available for all video sequences in this dataset.

- **Floorball Selection**
  Two video sequences created from the floorball data described in previous section for the purpose of evaluating the ghost detection method we propose. Groundtruth is available for all frames.

3 Background Subtraction for Object Tracking

3.1 Introduction

In many surveillance applications, background subtraction is an important first step. It is used to separate the possible interesting moving objects - *the foreground*, from the uninteresting part of the input image - *the background*.

The simplest method to achieve this goal is to compare every frame of the input video sequence with a static background image. However, having a single constant image to represent the background proves insufficient in most real-world applications. Sometimes the background image is not available at all. The background can contain unimportant moving objects like water and trees, the lighting conditions may vary or the scene can be permanently changed when objects are added or removed. To handle these issues, it is necessary to maintain a background model which can be updated with new information.

There are many different techniques that were proposed to model the background [6]. Modelling the value of each pixel by a mixture of Gaussians is a very popular method and provides state of the art results [7][8].

We focus on background subtraction as a preprocessing step for object tracking. It is common that an object stays relatively still for longer periods of time - long enough for the background model to accept it as a background. A resumed movement results in a false detection on the previous location of the motionless object. The detection does not correspond to any real object. We call this type of false detection a ghost. The same situation occurs when there are moving objects present at the time of the background initialization. These issues are caused by the background modelling and occur in many background subtraction algorithms. Here we propose a method to identify and avoid these false detections. An article where we describe the method was also accepted for Poster 2015, an international student conference at FEE, CTU in Prague.

3.2 State of the Art

Dealing with the problem of ghost detections is common for many background subtraction algorithms. General overview of the topic is given in [9].

A popular solution, which is convenient for many applications, assumes that every detection which does not move during multiple frames is a ghost [10][11][12][13]. We explain why this approach is not suitable for our purpose.

A method presented in [14] utilizes temporal image analysis and uses currently invisible background pixels in the background model update. In another work, [15], the ghost detection elimination is achieved by examining a pixel neighbourhood during the evaluation process and is built-in in the background subtraction process.

In [16], the authors combine gradient of pixel intensity and pixel color. Gradient difference is used in [17] to validate object detections. We take a similar but simpler approach.

3.3 Proposed Approach

First, it is important to define what is considered a ghost. In some of the works dealing with ghost detections, any detection that does not move in a defined period of time is labelled as a ghost and the background model is updated to accept it as background immediatelly. But it is common in tracking applications that some objects of interest do not move significantly for a long time so it is undesirable to label all stationary detections as ghosts.

Our approach is based on the same idea as Javed et al. [17]. In an image, every real object is separated from its surroundings by an edge. If an object does not have an edge on its border, it is not a real object but a ghost detection. In this sense every stationary detection that does not correspond to any real object in the image is considered a ghost. As an alternative to this edge based ghost detection we also propose a color based ghost detection. Ghost detections are identified using color histograms of inside and outside regions of the object.

The entire process can be split into three parts: background subtraction, ghost detection and background model update. The background subtraction can be any method that works with pixel by pixel model of the background. In our case the pixels are modelled by a mixture of Gaussians [7]. The output is a foreground binary mask.

The ghost detection step splits the foreground mask into single connected components, each of them being considered one object. Contours of objects are evaluated for a presence of image edges, or based on the color, and labelled as valid detections or ghosts.

The background model update step takes all the pixels in all the connected components classified as ghost detections and updates the corresponding pixels in the background model to the current value of those pixels. This way, ghost detections are quickly absorbed into the background.

The details of each step follow.

Table 1: Notation

I	current image
BG	background model
B	background image
$\mathcal{G}_i^x$	i-th gaussian modelling value of x
N^x	number of gaussians used to model value of x
F	foreground mask
α	learning coefficient of the background model
C_i	i-th connected component in current F
C_i'	expanded border of C_i
$E_{C_i}^I$	edge mask of I in region C_i
G	ghost mask
H_i^r	color histogram of the region C_i^r

3.3.1 Background Subtraction

The proposed method can be used with any background subtraction algorithm that maintains an independent model for each pixel. Let us denote I an image being processed, x being the image pixel and $I(x)$ the pixel value. Here, we use Gaussian mixture model. The detailed description can be found in [7], we present a short overview below.

Let us denote the background model BG, then $BG(x)$ is the model of pixel x, three Gaussians $\mathcal{G}_i^x$, $i = 1, 2, 3$, are used to model its value. μ_i^x is the mean of $\mathcal{G}_i^x$, $\Sigma_i^x = \sigma_i^{x2}\mathbf{1}$ its covariance and w_i^x its weight in the model.

To derive the background pixel value the Gaussians are first ordered by the fitness value w_i^x/σ_i^x and then the first N^x are used to compute the expected pixel value.

$$N^x = \arg\min_n \left(\sum_{i=1}^{n} w_i^x > T\right), \tag{1}$$

where the threshold T is the minimum prior probability of the background in the scene.

We evaluate the pixels to create a foreground mask F, where $F(x)$ is its value at pixel x. $F(x) = 1$ classifies the pixel x as a foreground, $F(x) = 0$ as a background.

A pixel x is considered as a background if there is a Gaussian in the model which matches the pixel value. We say that a Gaussian matches pixel value if its intensity is closer than 2.5 standard deviations from the Gaussian mean:

$$F(x) = \begin{cases} 1 & \text{if } \exists i < N^x : |I(x) - \mu_i^x| < 2.5 \cdot \sigma_i^x \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

The matched Gaussian parameters are updated as follows:

$$\begin{aligned} w_i^x(t+1) &= w_i^x(t) \cdot (1-\alpha) + \alpha, & (3) \\ \mu_i^x(t+1) &= \mu_i^x(t) \cdot (1-\alpha) + \alpha I(x), & (4) \\ \Sigma_i^x(t+1) &= \Sigma_i^x(t) \cdot (1-\alpha) + \alpha(I(x) - \mu_i^x(t))^2, & (5) \end{aligned}$$

where $\alpha \in \langle 0, 1\rangle$ is a learning coefficient which determines how fast the background model reacts to a change. If there is no matching Gaussian in the model, the Gaussian with the lowest fitness is replaced. The $I(x)$ is used as the new mean with a high value for variance and a low weight.

At $t = 0$, when the first frame is processed, the situation is same as if the learning coefficient was selected to $\alpha = 1$. The second frame at $t = 1$ provides 50% of the available information to the background model, however according to the rules above has a smaller effect, depending on the value of the learning coefficient α.

To deal with this problem we propose to use a modified learning coefficient α', with value $\alpha' = 1$ at $t = 0$. α' is gradually lowered until the desired value α is reached:

$$\alpha' = \max(\alpha, \frac{1}{1+t}) \tag{6}$$

This approach better reflects information value of the new frames. The difference between α and α' is illustrated in Figure 1.

For the purposes of background subtraction we decided to use CIE LUV color space, because the color difference better corresponds to the human perceptual difference [18].

This background model can handle both sudden and gradual changes in illumination, moving objects in the background and also permanent changes in the scene.

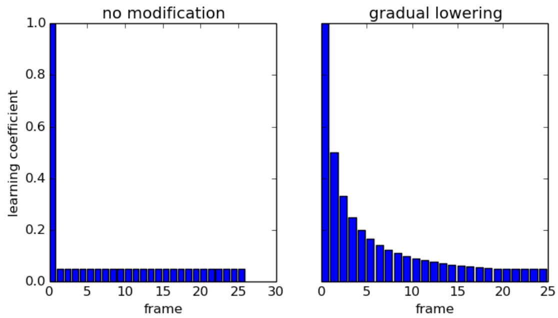


Figure 1: At the beginning of an image sequence the background model has to be initialized. When a background image without moving objects is not available, the model has to be learned from the sequence itself. On the left subfigure the learning coefficient $\alpha = 1$ for the first frame. The background model is completely initialized with the first frame. On the right subfigure the learning coefficient is smoothly decreased from 1 to the predefined value.

3.3.2 Edge Based Ghost Detection

This step takes a foreground mask F produced in the background subtraction step and evaluates validity of the detections based on edge presence on the borders of objects. First the connected components are extracted from the foreground mask, each is then evaluated independently.

$$F = C_1 \vee C_2 \vee \cdots \vee C_n \tag{7}$$

where $C_{i \in 1..n}$ is a single connected component. Only C_i formed by larger number of pixels than a size threshold T_S continues to be evaluated.

The border of C_i is denoted C_i'. Because the region C_i' will be used for edge detection, it needs to be wider than one pixel, so it is expanded as illustrated in Figure 2. Canny edge detector with Otsu threshold is used to compute the edges. By using the Otsu algorithm, we avoid selecting the threshold value manually.

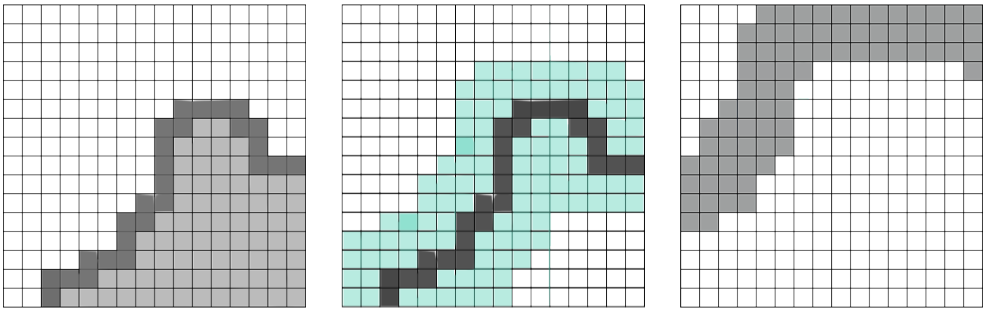


Figure 2: The left figure shows a part of an object C_i: it's border (dark grey) and interior (light grey). In the centre, there is the expanded object border C_i' (both light blue and dark grey), true border C_i^{tb} (dark grey) and its expansion C_i^{eb} (light blue). On the right figure the outer object border C_i^{ob} is displayed. It is the region adjacent to the expanded object border.

Edge mask obtained from currently processed image I in the region C_i' is denoted $E^I_{C_i'}(x)$. Image I contains the moving objects as well as the background so it follows that edges E^I from any region contain edges of the moving objects and also background edges. We are only interested in the edges belonging to the moving objects, which we can get by subtracting the background edges. The background image B is reconstructed from the background model, described in previous section, by:

$$\forall x \in B : B(x) = \mu_{k_m}, \quad k_m = \arg\max_k (w_k/\sigma_k). \tag{8}$$

Note that k_m is an index of a Gaussian with the highest fitness value w_k/σ_k.

Edge mask computed using the background image B in the region C_i' is denoted $E^B_{C_i'}$, similar to the case of I. By combining the image edges $E^I_{C_i'}$ with the background edges $E^B_{C_i'}$ we get the foreground edges:

$$E^F_{C_i'} = E^I_{C_i'} \wedge \neg E^B_{C_i'}. \tag{9}$$

Edges belong to the foreground objects if they are present in the current image I but not in the background image B.

By performing this evaluation only on the detected objects, it is possible to increase efficiency by computing the edges in small regions corresponding to the object's boundary instead of the whole image.

We propose several criteria functions that can be used to determine if object C_i is a valid detection or not. Object C_i is valid detection if the condition is satisfied. Output of this step is a binary ghost mask G.

- **Foreground edge probability:**

$$\frac{|E^F_{C'_i}|}{|C'_i|} > T_G, \tag{10}$$

  where $|X|$ is the number of (non-zero) pixels in X. The threshold T_G was empirically chosen to be 0.18.

- **Foreground edge ratio:**

$$\frac{|E^F_{C'_i}|}{|E^I_{C'_i}|} \cdot \max\left(\frac{|E^F_{C'_i}|}{|E^B_{C'_i}|}, 1\right) > T_G \tag{11}$$

  is a proportion of the foreground edges in all detected edges modified so that edges appearing in regions with no small number of background edges have higher value.

- **Narrow border:** As described above, for the purpose of edge detection the borders of objects were expanded. The region C'_i can be viewed as a conjunction of the one pixel wide true border region C^{tb}_i with the part obtained by the expansion C^{eb}_i, see Figure 2. Then using only edges on the true border C^{tb}_i:

$$\frac{\sum_{\forall x \in C^{tb}_i} E^F_{C'_i}}{\sum_{\forall x \in C^{tb}_i} E^B_{C'_i}} > T_G \tag{12}$$

  is a ratio of foreground edges to background edges on the true border of an object. We chose $T_G = 1$.

- **Edge probability difference:** We can select outer object border C_i^{ob}, as shown in Figure 2, so that C_i^{ob} and C_i' are disjunctive. By comparing probability of edges on the object border and outside border, we get a criteria function:

$$\frac{\frac{|E^I_{C_i'}|}{|C_i'|}}{\frac{|E^I_{C_i^{ob}}|}{|C_i^{ob}|}} > T_G. \tag{13}$$

  This criteria function has the advantage, that it does not require pixel by pixel background model.

3.3.3 Color Based Ghost Detection

The the advantage proposed color based ghost detection is that it does not require pixel by pixel background model. Connected components C_i in the foreground mask F are extracted the same way as for the edge based ghost detection (Equation 7) and thresholded by the same size threshold T_S.

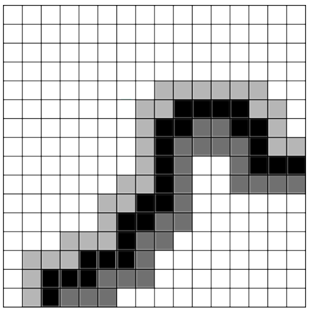


Figure 3: On the left figure is an object C_i: it's true border C_i^{tb} (black), inside contour C_i^{ic} (dark grey) and its outside contour C_i^{oc} (light grey).

True border of the object C_i^{tb}, inside object contour C_i^{ic} and outside object contour C_i^{oc} are shown in Figure 3. A color histogram of the current image I is calculated for each of these regions, that is H_i^{tb}, H_i^{ic} and H_i^{oc} respectivelly.

When the object C_i corresponds to correct detection, the color distribution inside the object, represented in H_i^{tb} and H_i^{ic}, should be similar and different from the color distribution outside of the object, H_i^{oc}. Correlation of histograms is used to compute the similarity s of two histograms:

$$s(H_1, H_2) = \frac{\sum_b (H_1(b) - \overline{H_1})(H_2(b) - \overline{H_2})}{\sqrt{\sum_b (H_1(b) - \overline{H_1})^2 \sum_b (H_2(b) - \overline{H_2})^2}} \tag{14}$$

where:

$$\overline{H_k} = \frac{1}{N} \sum_J H_K(J) \tag{15}$$

and N is the total number of histogram bins.

If the similarity of the outside contour region to the object border is different from the similarity of the two inside regions by more than a factor of given threshold T_G it is considered a ghost detection. We can write the criterium for the object C_i:

$$\left| \frac{s(H_i^{oc}, H_i^{tb})}{s(H_i^{ic}, H_i^{tb})} - 1 \right| > T_G \tag{16}$$

The value of the threshold was selected $T_G = 0.1$.

3.3.4 Background Model Update

The task of maintaining up-to-date background model is handled by the background subtraction algorithm described in step 1. We integrated the ghost detection results into the update process.

Taking the ghost mask G, all pixels x labelled as ghost detections are updated in the background model to the current value $I(x)$. The update is done by increasing the weight of the Gaussian $\mathcal{G}_i$ which matches the current value $I(x)$. When next frame is processed, $\mathcal{G}_i$ will be considered by the background model (Equation 1) and x correctly evaluated as background:

$$\forall x : \forall \mathcal{G}_i \in BG(x) : w_i^x = \begin{cases} \frac{w_i^x + \beta}{1+\beta} & \text{if } \mathcal{G}_i \text{ matches } I(x) \\ \frac{w_i^x}{1+\beta} & \text{otherwise,} \end{cases} \tag{17}$$

where β is a constant parameter. Following must also hold:

$$\forall x : \forall \mathcal{G}_i \in BG(x) : \sum w_i^x = 1. \tag{18}$$

3.4 Evaluation

To measure the performance of the proposed method, we compare the results obtained with different criteria functions and results without using the proposed ghost detection. Data used for the experiments are described in Sections 2.1 and 2.2.

Ideally, the results would be compared using the groundtruth as a reference. However, for PETS'09, floorball and parking sequences, background subtraction groundtruth is not available and the results are evaluated visually. PETS'09 and floorball sequences are used to evaluate the tracking system, our goal is to improve the background subtraction results on these sequences. The parking sequence was selected to demonstrate the effects of the method.

For Change Detection and floorball selection sequences the groundtruth is available, the results are evaluated using the following performance metrics (TP - True Positive, FP - False Positive, FN - False Negative, TN - True Negative):

- Recall:
$$Re = \frac{TP}{TP + FN}$$
- Precision:
$$Pr = \frac{TP}{TP + FP}$$

In Section 3.3.1, we use the CIE LUV color space for the background subtraction algorithm. We present a comparison to the results obtained with the RGB color space. Because we are interested in background subtraction for tracking applications the effects of ghost detection on the overall performance of the tracking system are also evaluated and results are presented in Section 4.7.

3.5 Experimental Results

Results of the proposed method on video sequence of parking car are presented in Figure 5. It shows that using the ghost detection eliminates false detection caused by movement of a car previously belonging to the background. Outputs of the foreground edge probability (Equation 10), narrow border (Equation 12) and color based (Equation 16) criteria functions are satisfactory. Foreground edge ratio (Equation 11) has almost no effect in this situation and edge probability difference (Equation 13) produces worse results than no ghost detection. We use only the first four criteria for the remaining evaluations.

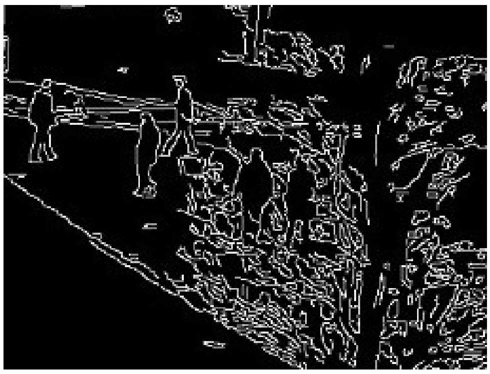
Figure 6 shows comparison of the 4 criteria functions on the floorball dataset. Foreground edge probability narrow border and color based criteria functions have again good results. The ghost artefacts are soon eliminated, and valid foreground objects remain in the foreground. The method works correctly on objects broken into several parts.

To simulate movement of objects that are considered background, we did not use the gradually changing learning coefficient at the initialization of the experiment on the PETS'09 dataset. The modification of the learning coefficient was described in Section 3.3.1. Figure 7 shows the results, which are similar to the floorball sequence.

The three criteria functions that performed well in previous experiments are evaluated using background subtraction groundtruth. For each data we present comparison of ghost detection results obtained using both CIE LUV and RGB color space. To simulate movement of objects that are considered background, we use the learning coefficient again and present the results with the original learning coefficient as well as the modified version.

Figure 8 shows the results on videos of intermittent object motion from Change Detection dataset. Narrow border criteria function using RGB color space performs same as the algorithm without ghost detection, remaining criteria offer significant improvement. The results on the entire Change Detection dataset are in Figure 9, all criteria functions perform similarly to the algorithm without ghost detection.

(a)



(b)

Figure 4: Illustration of edges present in the image background. (a) image from the sequence. (b) corresponding edge image

Results obtained on floorball selection data are presented in Figure 10. It shows significant improvement for all the ghost detection functions. The results on both floorball selection and Change Detection dataset are shown in Figure 11 confirming the possitive effect of all ghost detection criteria functions.

Foreground edge probability criterion function performs well in all the experiments described above and often provided the best result. Results of narrow border and color based ghost detection are also satisfactory.

The background subtraction results are improved in cases when there are moving objects present in the scene at the time of initialization, or when objects considered as background start to move. Ghost detections are eliminated and the background model is less affected by the initial values, which results in improved detections in those regions in the subsequent frames.

The proposed method, however, has some weak spots as well. Valid objects may be detected as ghosts when only part of the object is detected as a foreground or when the object is broken into parts. This can happen when objects, or their parts, have similar colors as their background or when there are background objects in front of them, e.g. tree branches or road signs.

A missing ghost detection may occur when it overlaps with a moving object. This may result in a delay in the ghosts elimination in case of very slowly moving objects. Missed ghost detections may also occur when there is a ghost detection in a region with complicated edges present in the background, as shown in Figure 4.

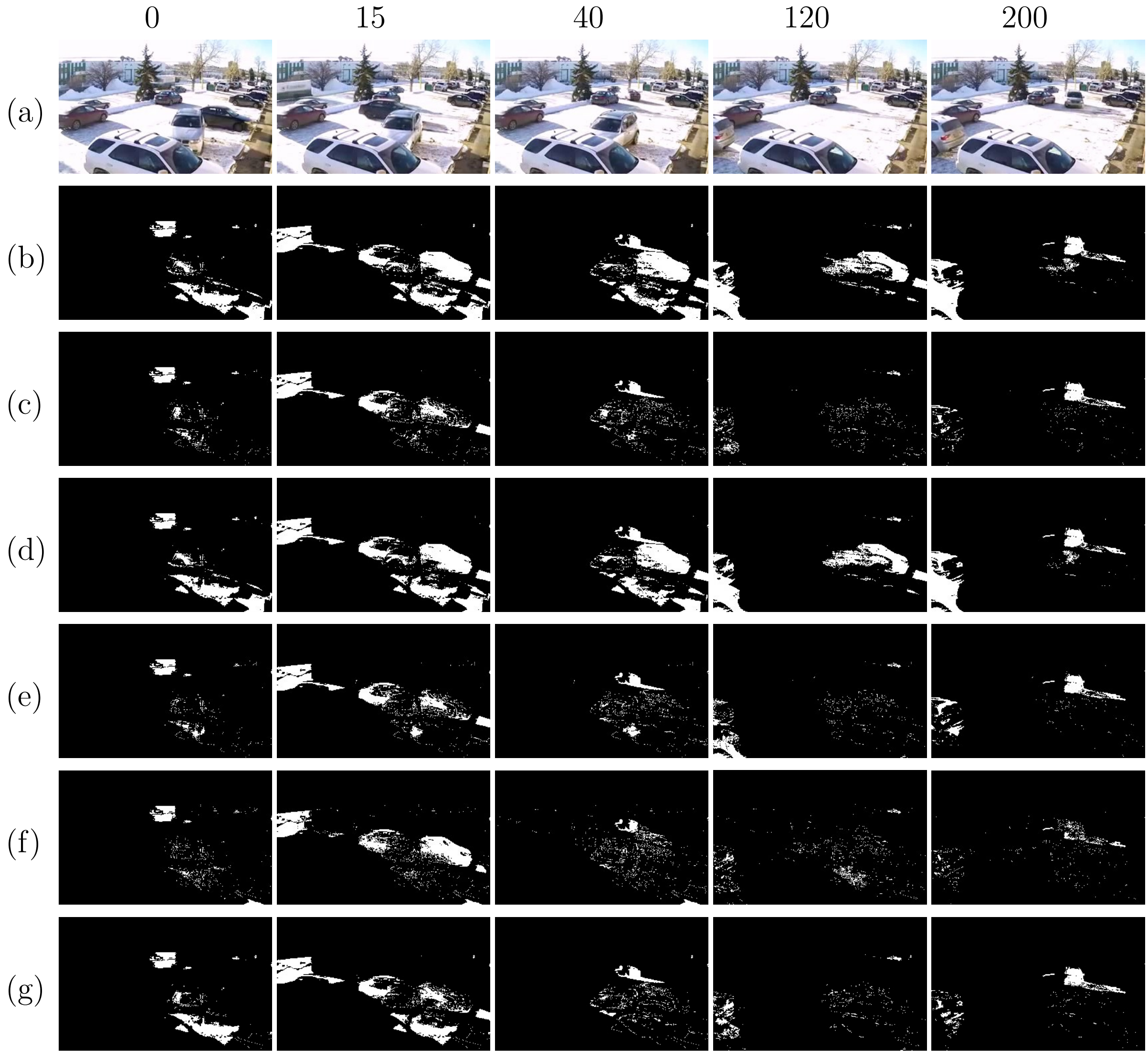


Figure 5: The background subtraction results of different criteria functions for ghost detection on the parking video sequence. The first row shows the frame number, frame 0 was selected from the middle of the sequence. (a) Images from the sequence. The rows bellow contain background subtraction results for different criteria functions: (b) No ghost detection. (c) Foreground edge probability. (d) Foreground edge ratio. (e) Narrow border. (f) Edge probability difference. (g) Color based ghost detection

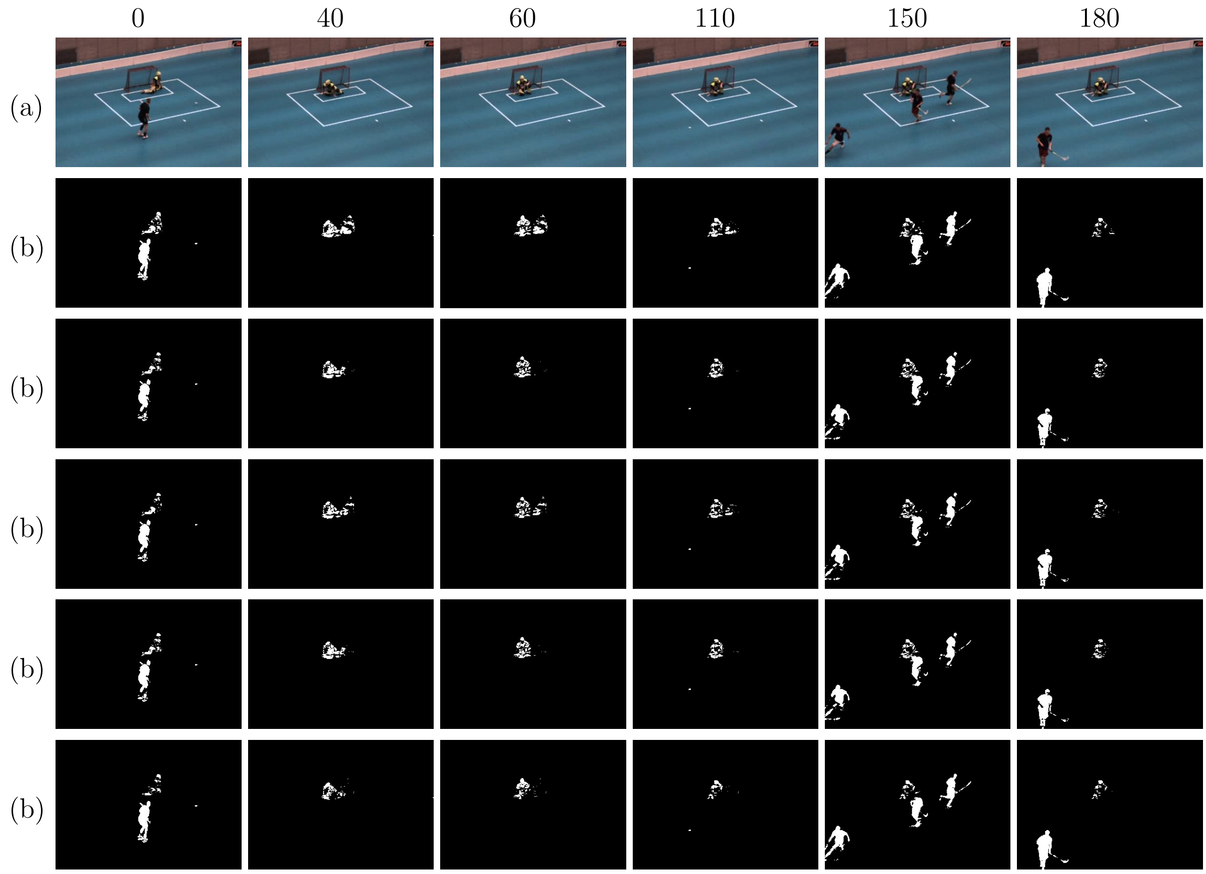


Figure 6: The background subtraction results of different criteria functions for ghost detection on the floorball video sequence. The first row shows the frame number, frame 0 was selected from the middle of the sequence. (a) Images from the sequence. The rows bellow contain background subtraction results for different criteria functions: (b) No ghost detection. (c) Foreground edge probability. (d) Foreground edge ratio. (e) Narrow border. (f) Edge probability difference. (g) Color based ghost detection

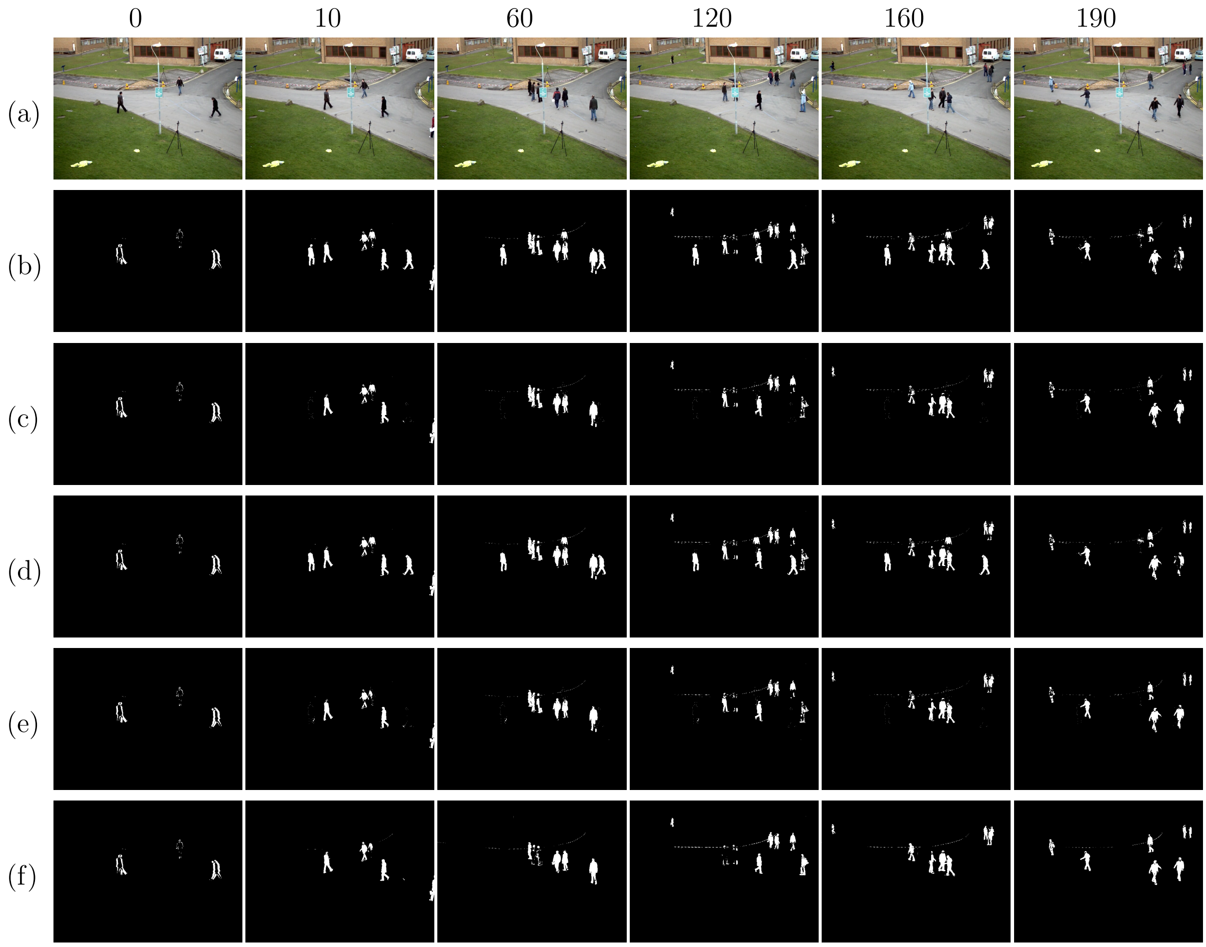


Figure 7: The background subtraction results of different criteria functions for ghost detection on the PETS'09 video sequence. For this experiment, the gradual change of the learning coefficient was not used in order to simulate situations, when objects in the background start to move. The first row shows the frame number, frame 0 is the first frame of the sequence. (a) Images from the sequence. The rows bellow contain background subtraction results for different criteria functions: (b) No ghost detection. (c) Foreground edge probability. (d) Narrow border. (e) Edge probability difference. (f) Color based ghost detection

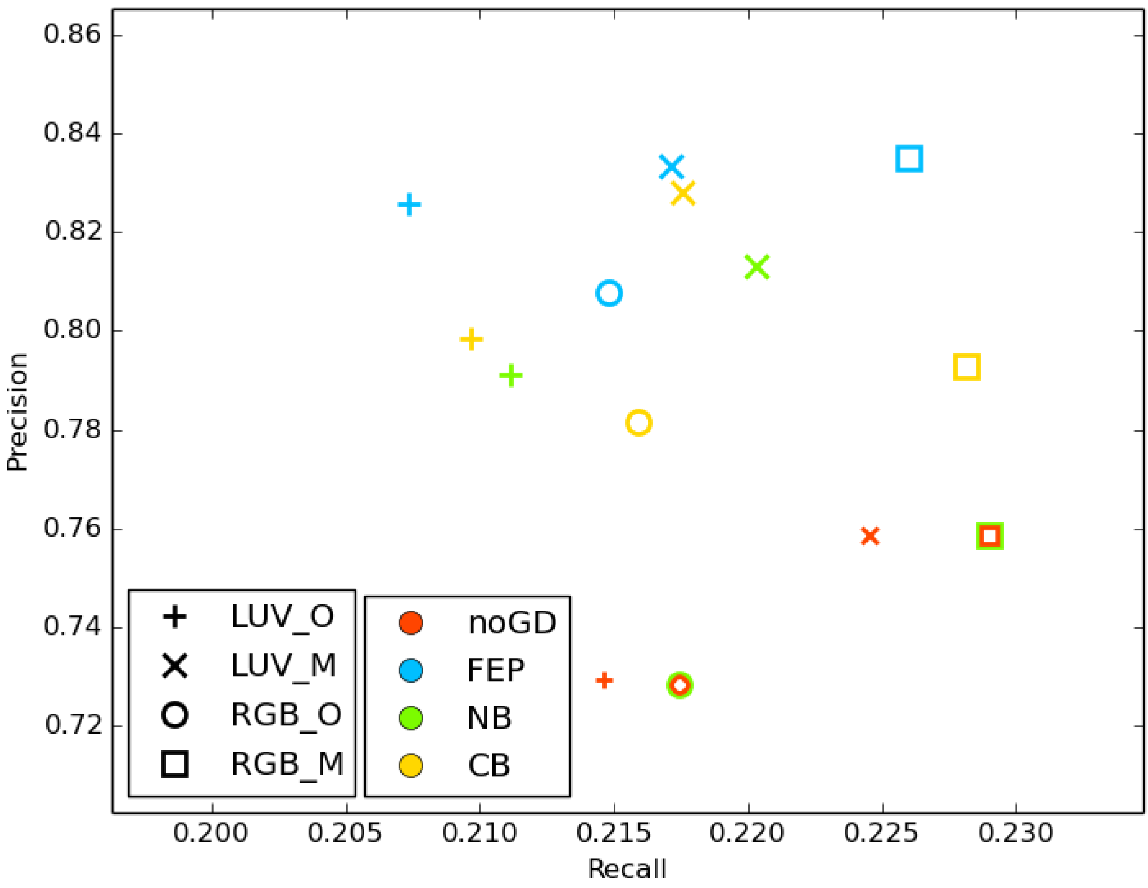


Figure 8: Evaluation of background subtraction using videos of intermittent object motion from Change Detection dataset. The figure compares background subtraction without ghost detection (noGD) and with ghost detection using different criteria functions: foreground edge probability (FEP), narrow border (NB) and color based (CB). Both the RGB and the CIE LUV color spaces are used with the original (O) and modified (M) learning coefficient at the initialization.

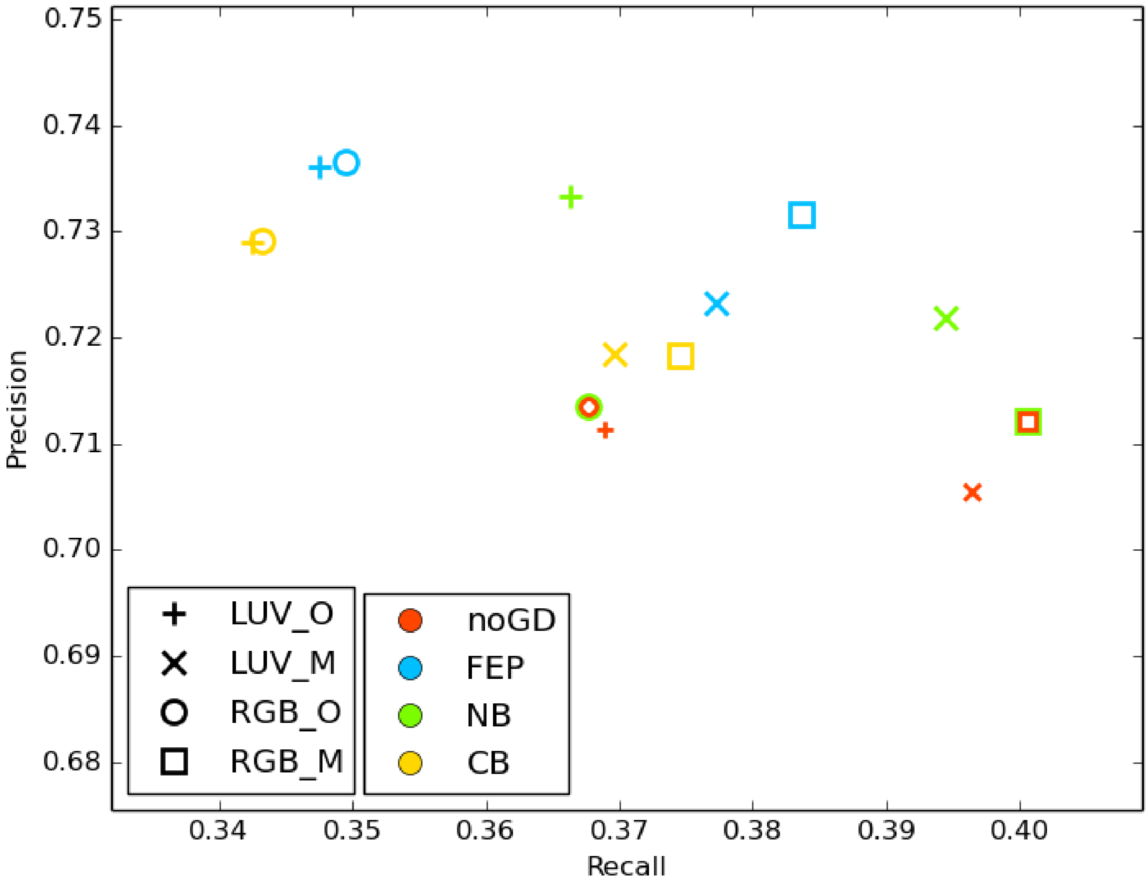


Figure 9: Same evaluation as on Figure 8 on the entire Change Detection dataset. That is evaluation of background subtraction without ghost detection (noGD) and with different methods for ghost detection: foreground edge probability (FEP), narrow border (NB) and color based (CB). Both the RGB and the CIE LUV color spaces are used with the original (O) and modified (M) learning coefficient at the initialization.

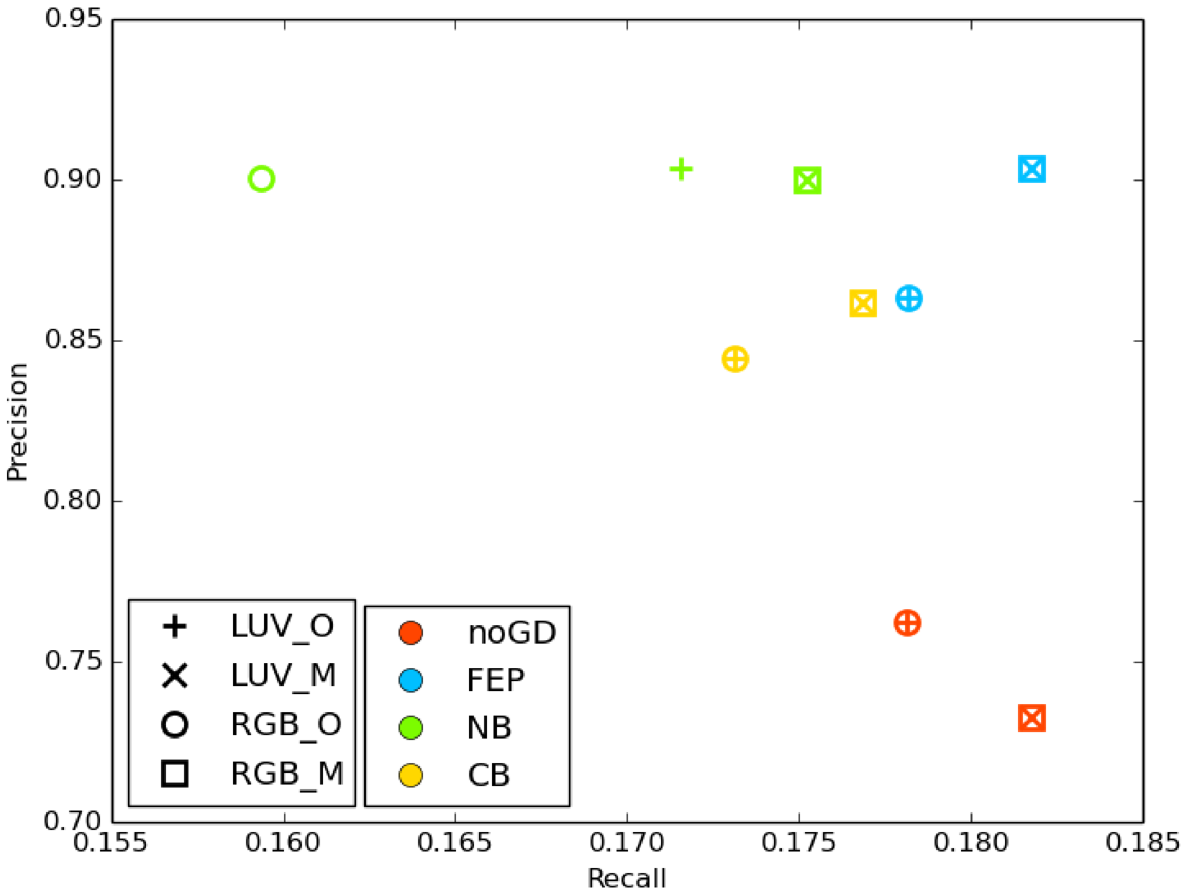


Figure 10: Same evaluation as on Figure 8 on floorball selection data. That is evaluation of background subtraction without ghost detection (noGD) and with different methods for ghost detection: foreground edge probability (FEP), narrow border (NB) and color based (CB). Both the RGB and the CIE LUV color spaces are used with the original (O) and modified (M) learning coefficient at the initialization.

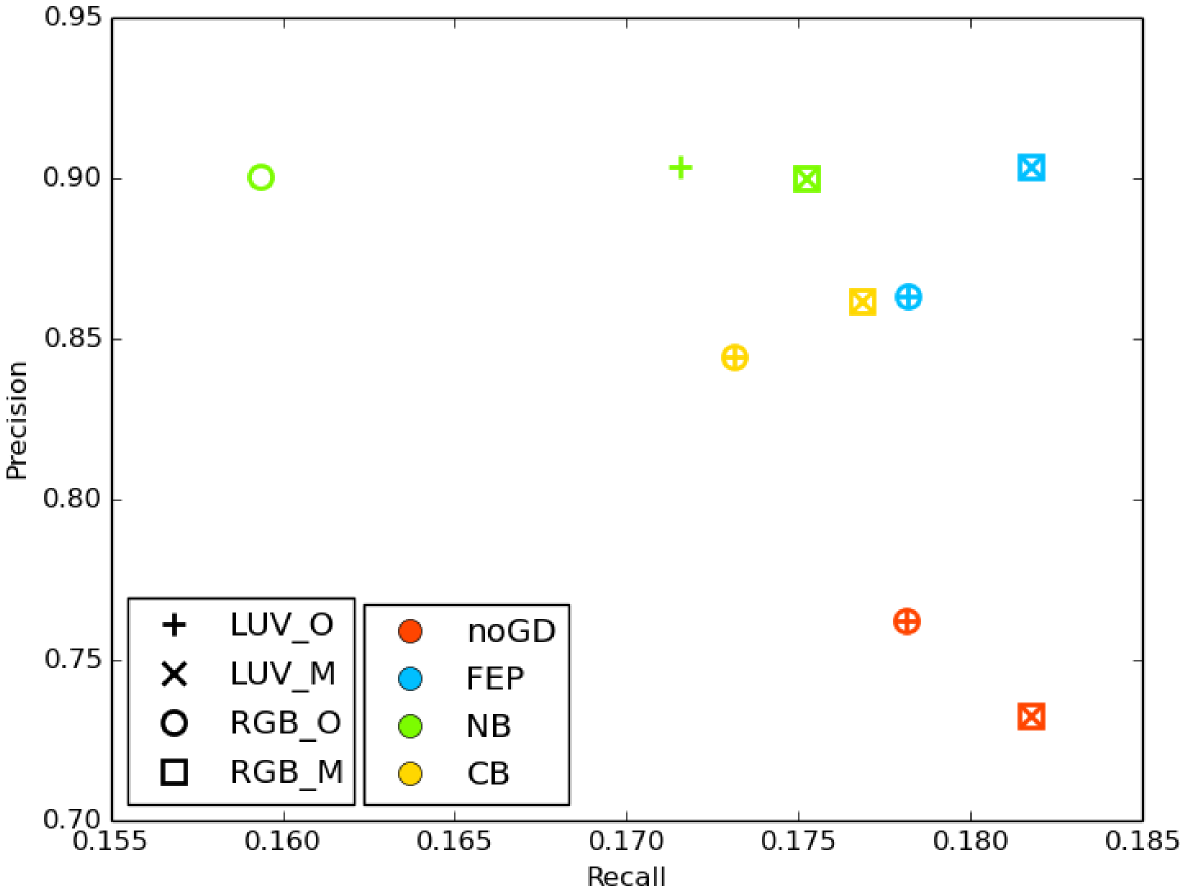


Figure 11: Same evaluation as on Figure 8 on both Change Detection and floorball selection data. That is evaluation of background subtraction without ghost detection (noGD) and with different methods for ghost detection: foreground edge probability (FEP), narrow border (NB) and color based (CB) on Change Detection dataset and floorball selection data. Both the RGB and the CIE LUV color spaces are used with the original (O) and modified (M) learning coefficient at the initialization.

4 Multi-View Multiple Object Tracking

4.1 Overview

This section is structured as follows. First we discuss ethical concerns regarding people tracking. Next we provide an overview of relevant previous work on the topic. The description of the tracking system itself and the modifications we propose can be found in Section 4.4. We continue by explaining the evaluation procedure of the tracking experiments followed by an experiment to determine the appropriate parameter settings. The last part, Section 4.7, presents all the tracking experiments conducted to evaluate the proposed modifications of the tracking system as well as the effect of ghost detection described in previous section on the system.

4.2 Ethical Consideration

Machines and technology in general became important part of our lives, most people use many different technologies every day. Devices or internet services often gather information about their users and there is so much information stored today in databases worldwide, that an individual has no way of controlling, or even knowing what information about themselves are accessible to others. This leads to rising awareness of the issues of privacy and its protection.

Surveillance camera systems are widely used in modern cities to monitor public spaces in order to prevent crime and assist the police in investigations. While some people feel more secure with the presence of security cameras, others consider it as an intrusion into their privacy. The impact of video surveillance on crime rate is still subject to research, but the results indicate the expected desirable effect[19]. However, the justification for such massive surveillance still remains part of a bigger question a modern society needs to answer, that is whether increased security is worth the loss of freedom that is inseparably linked with it.

Also, it is important to keep in mind that identity in the context of this thesis and in research on multiple object tracking in general is not a true identity of a person in the real world (eg. their name). Rather it is a unique arbitrary description chosen for the purposes of the tracking system. Preserving identity of an object means that during the entire tracking process this object is assigned the same description. After leaving and re-entering the scene the object receives a new identity. However, it should be noted that in many sport tracking applications the tracking system "is aware" of the true identities of the players and uses this information.

Automated monitoring systems, same as any other technology, can be used with bad intentions, but we believe that the possitive effects of its many possible applications exceed the dangers. It is important to carefully consider the benefits and potential threats before employing such a system. Extra attention should always be paid to fulfill all legal requirements and to avoid intruding into peoples lives any more than necessary.

4.3 State of the Art

Object tracking is a challenging problem and is the subject of extensive research [20] due to its importance in the field of computer vision. Our main focus is tracking of multiple objects for application in team sports and there are many different techniques used in this area [21]. The techniques can be classified as intrusive - they require special sensors to be placed on the tracked objects - or nonintrusive - for example vision systems. We aim for a general approach not limited to team sports, where special sensors are not needed.

Multiple cameras are often used for covering large areas or for complex scenes with common object occlusions. A method presented in [22] uses single view tracking and automatically switches cameras for the best available viewpoint. In [23], the tracking is performed in single cameras and then combined together to help resolve object occlusions. The approach uses constant velocity motion model and does not allow complex object motions.

A tracking system for indoor environments using RGBD cameras was presented in [24]. The volumetric information is used for background subtraction algorithm and different objects are identified based on the color information. A different system for people tracking in indoor scenes also based on RGBD cameras was described in [25]. Foreground detections are connected together to form people shaped clusters. Color histograms are used to preserve object identities. This system was designed to be used for a smart environment and evaluated in practical experiment.

An approach based on using agents to monitor and describe the scene proposed in [26] can be used for tracking but also for understanding the scene and different object interactions.

An efficient method for handling occlusions in a crowded scene was introduced in [27]. Object movement is limited to the ground plane and tracking is performed by combining all camera inputs. The method can accurately segment detected foreground components into corresponding objects. Another method that deals with occlusions in crowded scenes was presented in [28]. The occlusions are resolved by tracking feet on the ground plane using information from the different cameras.

Probabilistic occupancy map people detector was proposed in [29]. It approximates object silhouettes with rectangles and iteratively estimates the occupancy probabilities using the information from all views. K-Shortest Paths algorithm is used to find trajectories. This approach was extended to preserve object identities in [30][31]. The tracking system we use in this thesis is based on this apprach. In [32], an extension of the method is presented, which is based on using radio-based localization for identity preservation. In contrast with the previous approach, this method relies on additional data to improve the system.

4.4 System Description

The tracking system is based on methods proposed in [29][30][31]. There are three modules in the system: background subtraction, probabilistic occupancy map computation and k-shortest paths optimization.

Diagram of the tracking system is shown in Figure 12. First the input video sequence images are processed by a background subtraction algorithm producing foreground binary images, which show the interesting moving objects. These are used as input of Probabilistic Occupancy Map algorithm that merges the information from all the available views and produces a localization map. K-Shortest Paths algorithm is then used to link locations in the localization maps from all the frames of the videoo sequence into different objects trajectories.

4.4.1 Background Subtraction

The background subtraction is a separate part of the tracking system and any suitable background subtraction algorithm may be used. If the tracker is to perform well, it is crucial that the background subtraction produces correct results, since all errors created at this stage propagate through the entire tracking process.

The problem of obtaining correct background subtraction results and minimizing the errors is discussed at length in Section 3.

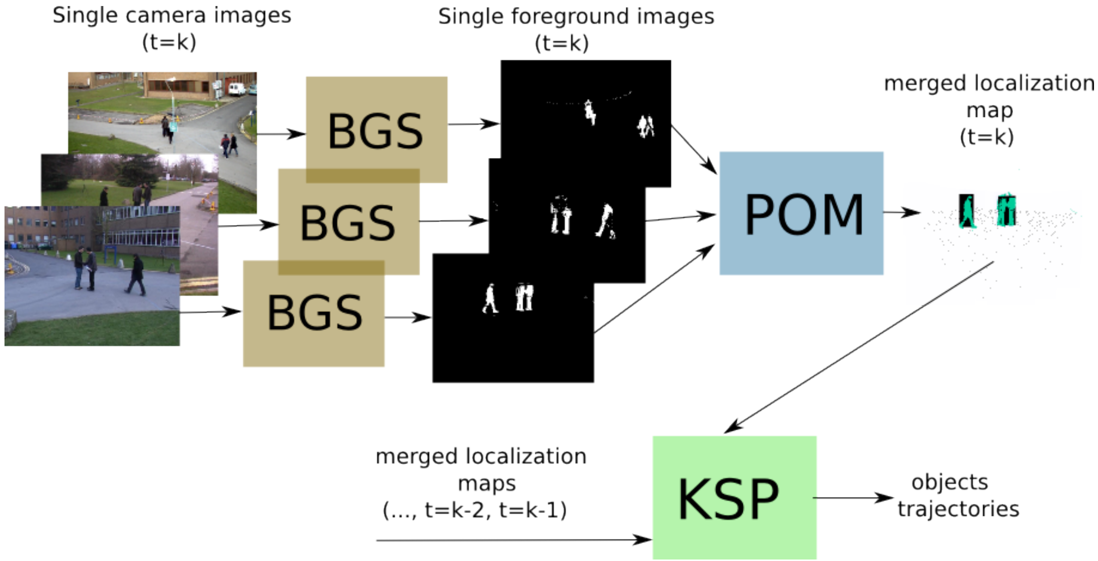
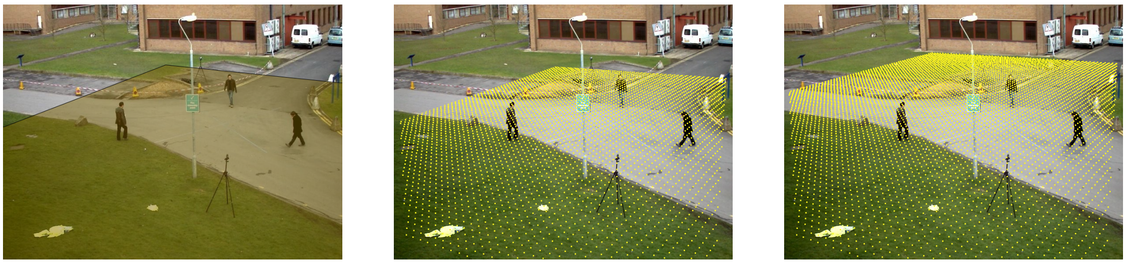


Figure 12: A diagram of the tracking system. Foreground binary images are extracted from the input images using background subtraction (BGS). Probabilistic Occupancy Map (POM) algorithm is used to merge the foreground images and to compute localization map. K-Shortest Paths (KSP) algorithm links the independant detections from multiple frames into optimal trajectories.

4.4.2 Probabilistic Occupancy Map

The Probabilistic Occupancy Map (POM) algorithm [29] is used to merge the information from all cameras and to compute the objects locations. The algorithm takes binary foreground images obtained by background subtraction as its input and produces the localization map.

The area of interest is partitioned into a regular grid of possible locations as displayed in Figures 13a and 13b. A simple appearance model is used to approximate object presence at each possible location with a rectangle. Dimensions of the rectangle are set as a parameter of the algorithm. For each frame, probability of occupancy of each cell in the grid is estimated using the available binary foreground images. Initially, all cells receive the same probability of occupancy, the estimated probabilistic map of occupancy is then projected into the input binary foreground images and iteratively optimized to achieve minimal difference from the input images.



(a) (b) (c)

Figure 13: (a) An image from the PETS'09 sequence with highlighted area of interest where the tracking takes place. (b) The same image with the area of interest partitioned into a discrete grid. (c) The same image showing the extended area of interest.

By analyzing the POM results in situations where the tracking system showed incorrect behavior, we identified two weak points:

- objects outside the area of interest
- objects on the edge of field of view

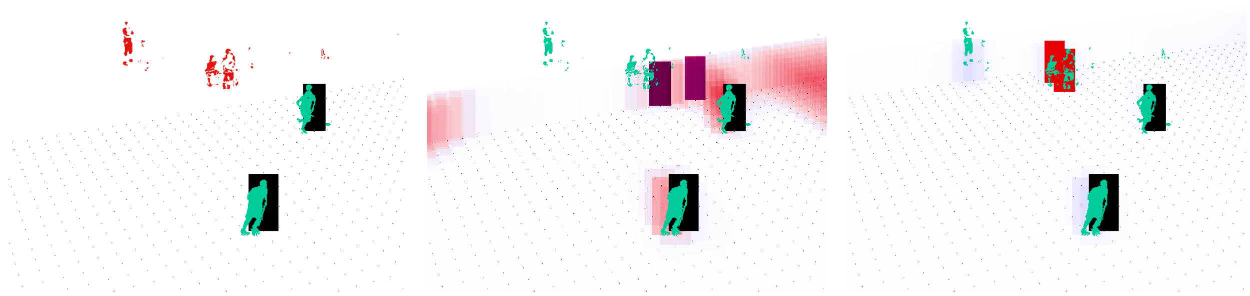
Objects Outside of the Area of Interest

The area of interest is partitioned into a discrete grid of possible object locations. The POM is computed based on intersection of the object appearance model, a rectangle, placed on every possible location in the grid with the binary image produced by the background subtraction. It is common that the field of view of the given camera contains part of a scene outside of the area of interest, where objects may be present. Information about these objects provided by the background subtraction is not used to improve POM inside the area of interest. Furthermore, depending on the camera's field of view, objects outside of the area of interest often partially intersect with the object appereance model in possible locations leading to deterioration of the POM accuracy. In extreme cases, it may cause false detections or identity switches when an object leaves the area of interest at the same time as another object enters in approximately the same location.

We propose to extend the area of interest for the computation of POM, as illustrated in Figures 13b and 13c. Therefore the correct location is estimated by the POM algorithm for objects outside of the area of interest, but inside its extension, and their impact on the result inside the area of interest is minimized. There may still be objects present even outside the extended area. The size of the extended area should be selected in such a way that it will contain the effects caused by outside objects and preserve accurate POM inside the area of interest.

However, with wider area the time needed to compute the POM rises, so the extended area should be selected reasonably, keeping in mind the current scene dispositions and the computational requirements. An automatic computation for the optimal size of the extension could be based on the angle between the camera's optical axis and the ground plane.

Figure 14 shows the result of this solution. In Figure 14a is the ideal POM result when an object is present outside of the area of interest, the POM should reflect only objects inside the area of interest. However, the object outside of it has significant effect on the POM accuracy as can be seen in Figure 14b. Figure 14c shows the result with extending the area of interest as described above, in Figure 13c is the extended area of interest in the original image.



(a) **(b)** **(c)**

Figure 14: POM visualization of a situation when an object is present outside of the area of interest: (a) Ideal result. The outside objects are highlighted in red. (b) Real result. The raised probability of occupancy caused by the outside objects is highlighted in red. (c) Real result with extended area of interest, the detections of outside objects are highlighted in red.

Objects on the Edge of the Acquired Image

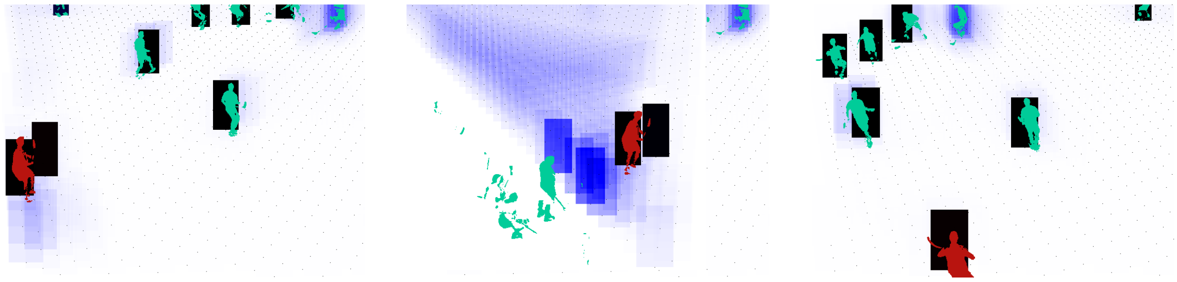
When the field of view of a camera does not contain the entire area of interest, the possible object locations are considered only if they are visible in the particular view. An object location should be considered visible if an object present at that location is visible in the image acquired from the camera. There may be different rules to decide about the visibility of an object location, two different simple rules are:

- Ground visibility - possible object location is visible in a camera when the point on the ground plane corresponding to that location is visible
- Complete visibility - possible object location is visible in a camera when the entire object appereance model placed at that location is visible

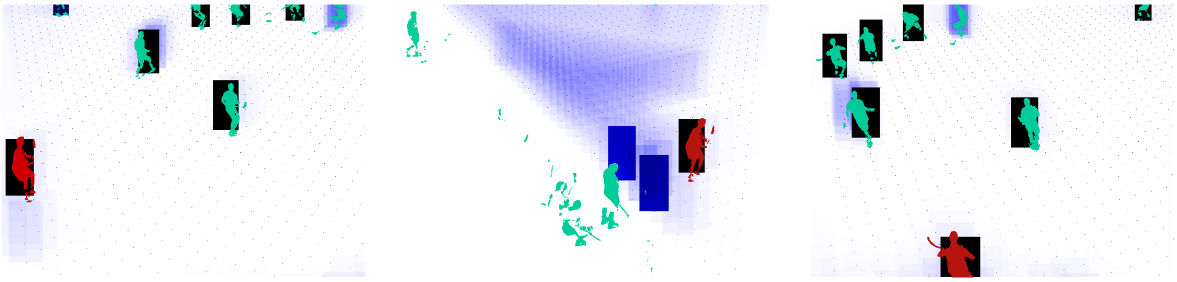
The ground visibility rule is inconvenient because it favors the ground plane, it is vertically assymetrical and completely ignores the height of an object. Using the complete visibility rule, we found that it is too restrictive for objects on the edges of the acquired image, disregards the available information and often causes double detections, as shown in Figure 15a.

The ideal solution is using continuous visibility value. Every possible object location would then have visibility value assigned, equal to the proportion of the appearance model that is visible. POM would be computed using the visibility value as weight of each location to reflect its informative value. However, implementing this solution would require extensive modifications of the POM algorithm, making it more complex and increasing the computation time.

We propose a center of gravity rule. The center of gravity of the object appearance model is computed for each possible location. Visibility of each location is determined by the visibility of the center of gravity of the corresponding object model. In practice it means that more than 25% of the object is visible for objects in the corners of the acquired image and more than 50% for other objects on its edge. The result is shown in Figure 15b.



(a)



(b)

Figure 15: POM visualization of a situation when an object (highlighted in red color) is present on the edge of the acquired image. Images taken from different cameras: (a) Double detection caused by complete visibility rule. (b) Correct result using center of gravity rule.

4.4.3 K-Shortest Paths Optimization

K-Shortest Paths algorithm is used to form optimal trajectories based on the POM results from different frames.

The algorithm finds k paths corresponding to single object trajectories inside a graph formed from the grid nodes connected over a sequence of frames. The cost of edge between two nodes depends on the probability of occupancy of the start node. The total cost of the paths is minimized.

To account for objects entering and leaving the area of interest, the algorithm uses two virtual nodes - source and sink. These virtual nodes are connected to other nodes in the graph that correspond to locations where objects may enter or leave the area of interest. Only paths starting in the source node and ending in the sink node are considered. This way, no flow can be created or suppressed anywhere in the graph, except at the virtual nodes, no object can appear out of nowhere or disappear inside the area of interest.

Because at most one object can be present at any single location at given time, the order in which the nodes are assigned to different paths can have significant impact on the result. The paths are ordered based on their cost and they are processed from the lowest cost to highest. This solution gives priority to paths with low cost over paths with higher cost and insures that clear trajectories will not be disrupted.

4.5 Evaluation

The performance of the tracking system is evaluated using datasets described in Section 2.1. We decided to use CLEAR MOT[33] metrics, a standard evaluation tool widely accepted by the tracking community. The metrics uses two measures to describe a tracker's performance, multiple object tracking accuracy and multiple object tracking precision.

MOTA (multiple object tracking accuracy) reflects the number of errors that occur during the tracking process. Well performing tracker maximizes the MOTA value. It is defined as:

$$MOTA = 1 - \frac{\sum_t (FP(t) + FN(t) + ID(t))}{\sum_t NT(t)}$$

where $FP(t)$ is the number of false positives, $FN(t)$ the number of false negatives, $ID(t)$ the number of identity switches and $N(t)$ the number of objects present at time t.

MOTP (multiple object tracking precision) shows the ability of the tracker to estimate precise object positions. It is the average distance of the estimated positions to the ground truth positions, lower values mean the positions estimated by the tracker are accurate. MOTP is defined by:

$$MOTP = \frac{\sum_{t}, id(D_i^t), GT_i^t)}{\sum_t NM(t)}$$

where D_i^t is a detected target hypothesis, GT_i^t the groundtruth target and $NM(t)$ the number of matches found at time t.

To analyze the results and identify the critical situations, we evaluate the results visually using a visualization tool described in Section 5.2 and images generated in each step of the tracking proccess, that is foreground binary images and POM visualization.

4.6 Parameter Settings

For the purpose of POM computation the area of interest is partitioned into a discrete grid. The resolution of the grid is determined by the size of a single cell, which is set using a cell_size parameter.

A series of experiments was conducted on the floorball dataset to evaluate the effect of different values of the cell_size on the overall tracking system performance. Values in range from 0.3 m to 3 m were used in the experiments and the results are displayed in Figure 16. The highlighted value, cell_size=0.5 m, is used for further experiments, because it provides satisfactory results and it is easy to compute the location coordinates. For tracking on PETS'09 dataset, value cell_size=0.3 m was used because the area of interest is smaller compared to the floorball dataset.

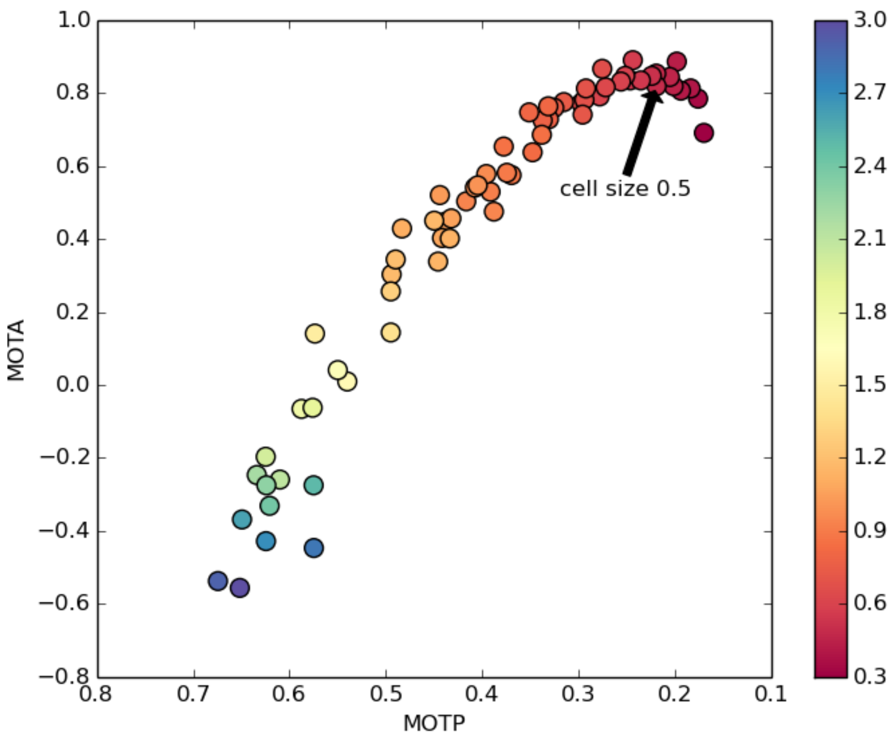


Figure 16: Overall tracking performance for cell_size parameter values in range from 0.3 m to 3 m, the values are coded by color corresponding to the color bar on the right. The marked value is used in the subsequent experiments.

4.7 Experimental Results

To evaluate the effect of ghost detection on the overall performance of our tracking system, we selected foreground edge probability criteria function (Equation 10). This function provided good results in sense of correctly identifying most ghost detections while it often allowed small and only partially detected objects to remain as foreground.

In Section 3.3.4 we use the CIE LUV color space for the background subtraction algorithm. We present a comparison to the results obtained with the RGB color space. By not using gradual lowering of the learning coefficient discussed in the same section, we simulate a scene where objects considered background start to move.

The tracking results[1] are presented in Table 2. We can see that the method offers a significant improvement on the performance of the tracking system. However, it has a slightly negative impact in scenes where it is rare for background objects to move.

The performance of the tracking system was evaluated using the proposed POM modifications. Figures 17 and 18 show the results for floorball and PETS'09 datasets, respectively. Details can be found in Tables 3 and 4. The tracker's performance is significantly improved by each modification, the improvements are similar for both datasets.

[1]Visualization is available at `http://1drv.ms/1APMorr` and on the included DVD

Table 2: This table summarizes the effects of ghost detection on the overall performance of the tracking system. The tracking is performed on 1000 frames of the floorball sequence and 794 frames of the PETS'09 sequence. Foreground edge probability criterion function (FEP, Equation 10) is compared with no ghost detection (noGD). In order to evaluate the influence of the color space selection in the background subtraction algorithm, both the CIE LUV and the RGB color spaces were used. No gradual lowering of the learning coefficient is used to simulate the movement of objects included in the background model.

Floorball sequences with gradual lowering of the learning coefficient						
Ghost Detection	Color Space	MOTA	MOTP	FN	FP	Identity Mismatches
noGD	RGB	0.7563	0.2284	66	5	7
FEP	RGB	0.7531	0.2298	66	5	8
FEP	LUV	0.750	0.2366	66	3	11
Floorball sequences without gradual lowering of the learning coefficient						
noGD	RGB	0.6031	0.2267	91	24	12
FEP	RGB	0.6718	0.2281	89	6	10
FEP	LUV	0.7063	0.2318	81	3	10
PETS'09 sequences with gradual lowering of the learning coefficient						
noGD	RGB	0.9016	8.4945	318	65	6
FEP	RGB	0.8973	8.5254	327	75	4
FEP	LUV	0.8933	8.3263	324	96	2
PETS'09 sequences without gradual lowering of the learning coefficient						
noGD	RGB	0.7191	9.0032	416	683	12
FEP	RGB	0.8890	8.4230	389	45	5
FEP	LUV	0.9019	8.4762	327	75	4

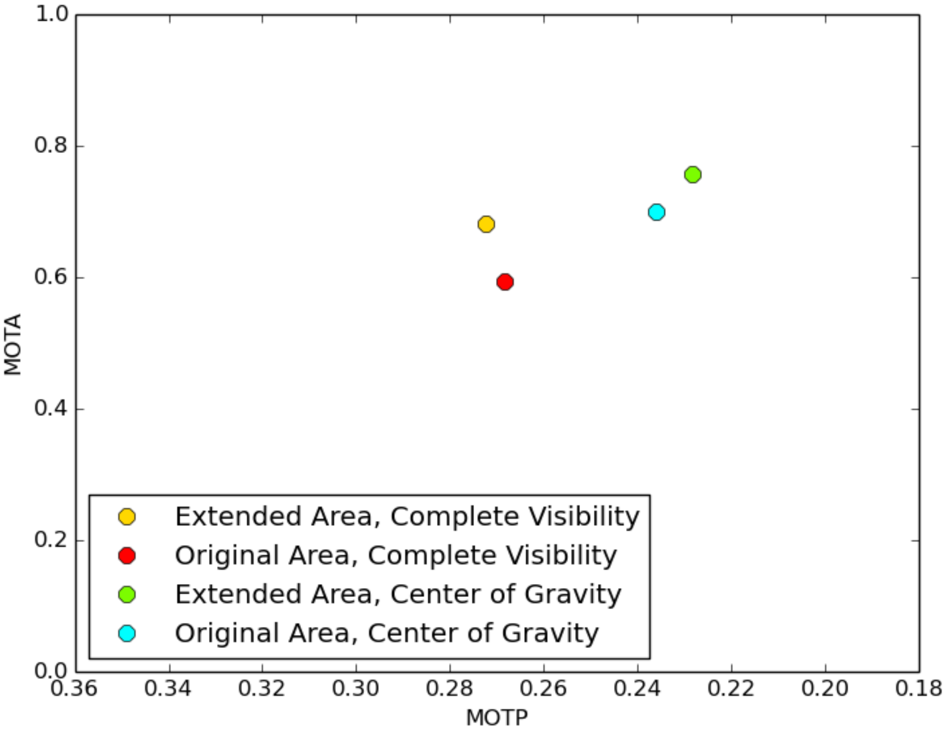


Figure 17: Tracking results on floorball data. The graph shows improvement of the tracker's performance for each POM modification.

Table 3: This table summarizes the effects of proposed POM modifications described in Section 4.4.2 on the overall performance of the tracking system. The tracking is performed on 1000 frames of the floorball sequences. The results are compared for tracking using the original size of the area of interest, the extended size of the area of interest, complete visibility rule and center of gravity visibility rule. Presented data show that both modifications significantly improve the tracking system performance.

Floorball sequence tracking results						
Area of Interest	Object Visibility	MOTA	MOTP	FN	FP	Identity Mismatches
Original Area	Complete	0.5938	0.2684	71	39	20
Original Area	Center of Gravity	0.7	0.2362	63	24	9
Extended Area	Complete	0.6813	0.2724	73	12	17
Extended Area	Center of Gravity	0.7563	0.2284	66	5	7

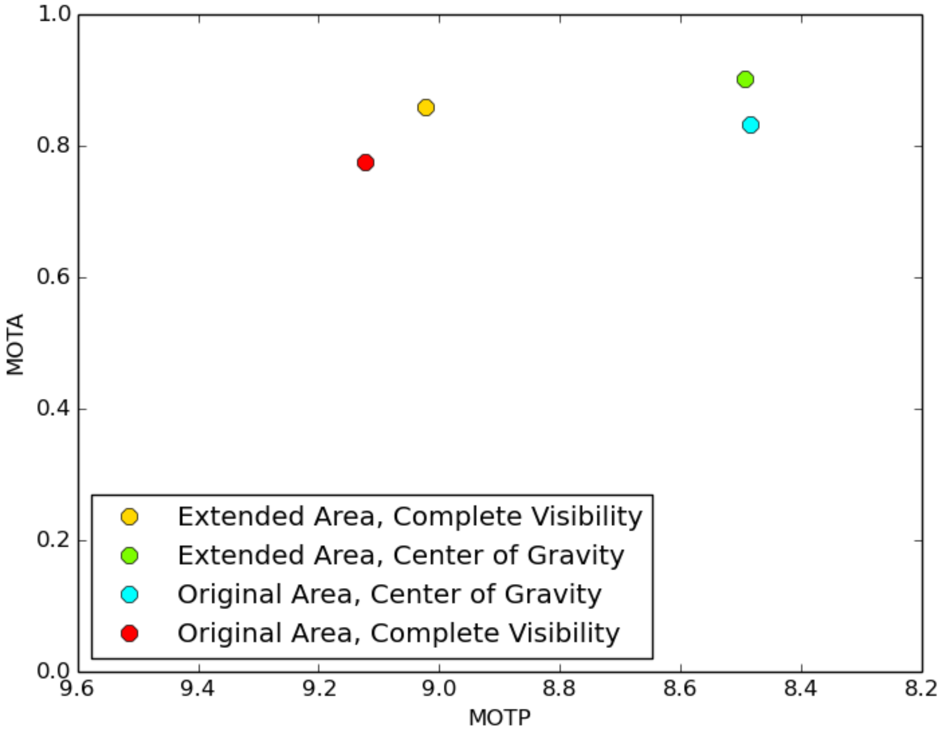


Figure 18: Tracking results on PETS'09 data. The graph shows improvement of the tracker's performance for each POM modification.

Table 4: This table summarizes the effects of proposed POM improvements described in Section 4.4.2 on the overall performance of the tracking system. The tracking is performed on 794 frames of the PETS'09 sequences. The results are compared for tracking using the original size of the area of interest, the extended size of the area of interest, complete visibility rule and center of gravity visibility rule. Presented data show that both modifications significantly improve the tracking system performance.

PETS'09 sequence tracking results						
Area of Interest	Object Visibility	MOTA	MOTP	FN	FP	Identity Mismatches
Original Area	Complete	0.7752	9.1230	189	691	9
Original Area	Center of Gravity	0.8329	8.4840	135	523	3
Extended Area	Complete	0.8587	9.0237	359	192	8
Extended Area	Center of Gravity	0.9016	8.4945	318	65	6

5 Supporting Tools

5.1 Background Subtraction Monitoring

Background subtraction monitoring tool was developed to facilitate the work on the ghost detection method. It allows observing the results of multiple background subtraction algorithms with different settings at the same time, as shown in Figure 19. It is possible to easily add or remove different views as well as change the parameters of the algorithms.

The single detections can be highlighted using different colors based on their properties. Multiple values that determine the function of the algorithm can be shown and linked with the corresponding detections for easier orientation. The images generated by this tool can be displayed immediatelly or saved as a video or single images for later use.

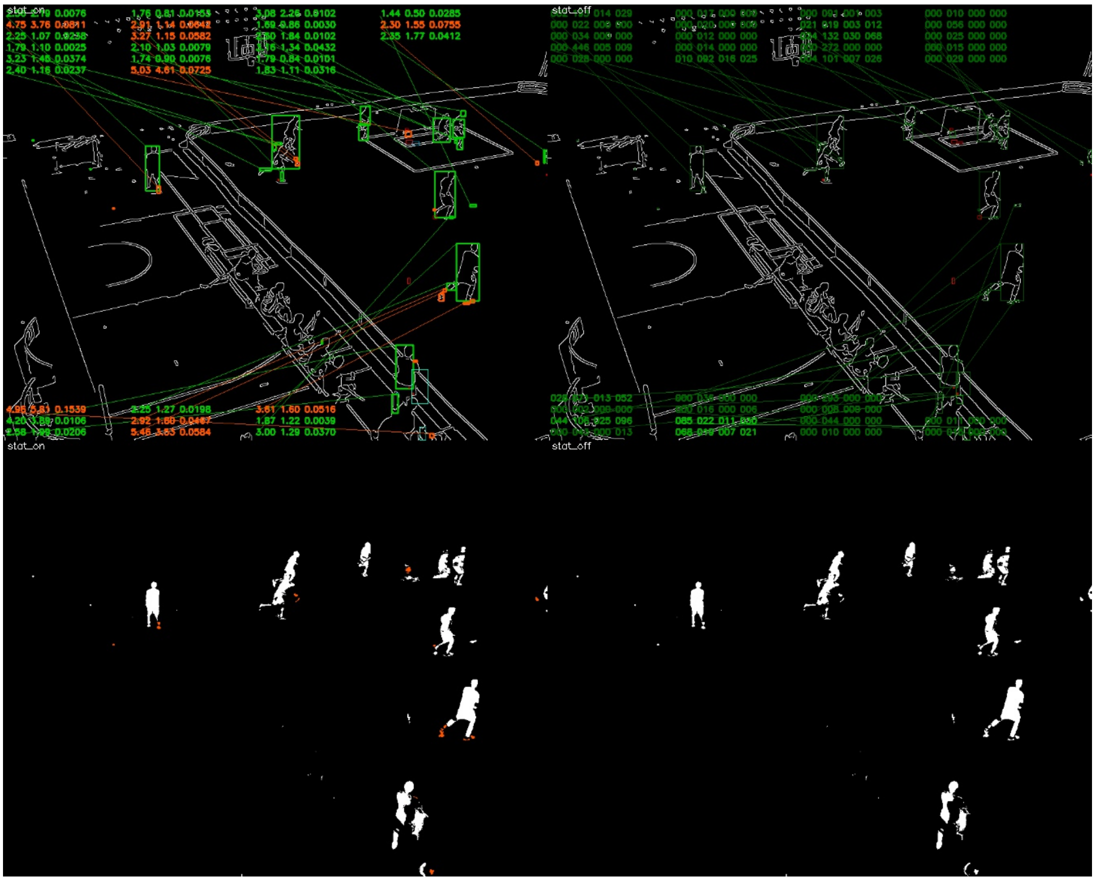


Figure 19: Background subtraction monitoring tool screenshot

5.2 Tracking Results Visualization

The purpose of tracking results visualization tool is to enable easy observation of the results of the tracking system. It works by loading required data from the tracking results, visualizing them, and saving as a video file. Screenshot of the video is displayed in Figure 20.

Images from all the cameras are shown on the right side, the images are automatically resized according to the number of cameras being used. The tracked objects are inside bounding boxes corresponding to the object location on the grid. The bounding boxes have different colors and are marked with the object's identity number. Past trajectories of the objects are highlighted. On the right side, there is a bird's-eye view projection of the scene. The area of interest for tracking is filled with the image of the playing field, its extension as described in Section 4.4.2 is shown in grey color. The number of each camera and the current frame are displayed in the top right corner of each view.

Figure 20: Tracking results visualization screenshot

6 Conclusions

First, we developed a general method for dealing with a special case of false detections - so called ghost detections - in background subtraction algorithms. The main advantages of the method are independence on a background subtraction algorithm and computational efficiency, because only small regions of the input image are evaluated.

Different criteria functions for the ghost detection were proposed, four of them based on detecting edges on objects' borders and one based on comparison of the color on the border of objects to the background.

Ghost detection method was evaluated visually and using background subtraction groundtruth on a dataset containing various indoor and outdoor video sequences. The results showed a positive impact on the background subtraction results. Foreground edge probability criterion function provided satisfactory results in all the experiments and often performed best of all the criteria functions.

The effect of ghost detection with foreground edge probability function on the overall performance of the tracking system was evaluated for indoor as well as outdoor datasets. We have incorporated the ghost detection into the background subtraction module of the tracking system and compared the performance with the original version. The method offers a significant improvement on the performance of the tracking system. However, it has a slightly negative impact in scenes where it is rare for background objects to move.

Then, we analyzed the output of the tracking system and identified situations in which errors occur most frequently. Those are situations when objects are present either outside of the area of interest for tracking or on the edge of the field of view of a camera.

In order to solve these issues, we proposed modifications of the probabilistic occupancy map module of the tracking system. The modifications consist of extending the area of interest and employing more suitable object visibility rule. The impact of the modifications on the tracking system's performance was again evaluated on both indoor and outdoor datasets, showing significant improvement.

The problems we faced are common for many automated tracking systems and the elaborated solutions can be successfully applied to improve them as well.

7 Future Work

Ghost detection deals with infrequently moving objects and eliminates the residue caused by a background object moving from its original position. The other side of the problem with infrequent object movement is objects that stop moving, because after a period of time they fade into the background. A method to detect objects of interest and protect them from becoming background would complement the ghost detection, together the two methods would offer a complete solution to infrequently moving objects in background subtraction.

The tracking system works by forming trajectories based only on the POM results, no information about the appearence of the objects being tracked is available. A possible next step in the work on improving the system is creating an object appearance model to describe the objects and using this information to help prevent identity switches.

References

[1] Anna Ellis, Ali Shahrokni, and James Michael Ferryman. Pets2009 and winter-pets 2009 results: A combined evaluation. In *Performance Evaluation of Tracking and Surveillance (PETS-Winter), 2009 Twelfth IEEE International Workshop on*, pages 1–8. IEEE, 2009.

[2] Anna Ellis, Ali Shahrokni, and James Michael Ferryman. Pets 2009 dataset. `http://www.cvg.reading.ac.uk/PETS2009/`, 2009.

[3] K. KIDZIDZI. Youtube video. `https://www.youtube.com/watch?v=HYcsW48HHXM`, 2014.

[4] N. Goyette, P. Jodoin, F. Porikli, J. Konrad, and P. Ishwar. Changedetection.net: A new change detection benchmark dataset. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, pages 1–8, June 2012.

[5] N. Goyette, P. Jodoin, F. Porikli, J. Konrad, and P. Ishwar. Change detection dataset. `www.changedetection.net`, 2012.

[6] Yannick Benezeth, P-M Jodoin, Bruno Emile, Hélene Laurent, and Christophe Rosenberger. Review and evaluation of commonly-implemented background subtraction algorithms. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4. IEEE, 2008.

[7] Pakorn KaewTraKulPong and Richard Bowden. An improved adaptive background mixture model for real-time tracking with shadow detection. In *Video-based surveillance systems*, pages 135–144. Springer, 2002.

[8] Zoran Zivkovic. Improved adaptive gaussian mixture model for background subtraction. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 2, pages 28–31. IEEE, 2004.

[9] S Cheung Sen-Ching and Chandrika Kamath. Robust techniques for background subtraction in urban traffic video. In *Electronic Imaging 2004*, pages 881–892. International Society for Optics and Photonics, 2004.

[10] Simone Calderara, Rudy Melli, Andrea Prati, and Rita Cucchiara. Reliable background suppression for complex scenes. In *Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks*, pages 211–214. ACM, 2006.

[11] Rita Cucchiara, Costantino Grana, Massimo Piccardi, and Andrea Prati. Detecting moving objects, ghosts, and shadows in video streams. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(10):1337–1342, 2003.

[12] Bijan Shoushtarian and Helmut E Bez. A practical adaptive approach for dynamic background subtraction using an invariant colour model and object tracking. *Pattern Recognition Letters*, 26(1):5–26, 2005.

[13] Hanzi Wang and David Suter. Background subtraction based on a robust consensus method. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 1, pages 223–226. IEEE, 2006.

[14] Paolo Spagnolo, M Leo, A Distante, et al. Moving object segmentation by background subtraction and temporal analysis. *Image and Vision Computing*, 24(5):411–423, 2006.

[15] Olivier Barnich and Marc Van Droogenbroeck. Vibe: A universal background subtraction algorithm for video sequences. *Image Processing, IEEE Transactions on*, 20(6):1709–1724, 2011.

[16] Sumer Jabri, Zoran Duric, Harry Wechsler, and Azriel Rosenfeld. Detection and location of people in video images using adaptive fusion of color and edge information. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, volume 4, pages 627–630. IEEE, 2000.

[17] Omar Javed, Khurram Shafique, and Mubarak Shah. A hierarchical approach to robust background subtraction using color and gradient information. In *Motion and Video Computing, 2002. Proceedings. Workshop on*, pages 22–27. IEEE, 2002.

[18] Janos Schanda. *Colorimetry: Understanding the CIE system.* John Wiley & Sons, 2007.

[19] Brandon Welsh and David Farrington. Effects of closed circuit television surveillance on crime: A systematic review. *Campbell Systematic Reviews*, 4(17), 2008.

[20] Alper Yilmaz, Omar Javed, and Mubarak Shah. Object tracking: A survey. *Acm computing surveys (CSUR)*, 38(4):13, 2006.

[21] Catarina B Santiago, Armando Sousa, Maria Luisa Estriga, Luis Paulo Reis, and Martin Lames. Survey on team tracking techniques applied to sports. In *Autonomous and Intelligent Systems (AIS), 2010 International Conference on*, pages 1–6. IEEE, 2010.

[22] Quin Cai and Jake K Aggarwal. Automatic tracking of human motion in indoor scenes across multiple synchronized video streams. In *Computer Vision, 1998. Sixth International Conference on*, pages 356–362. IEEE, 1998.

[23] James Black and Tim Ellis. Multi camera image tracking. *Image and Vision Computing*, 24(11):1256–1267, 2006.

[24] George Galanakis, Xenophon Zabulis, Panagiotis Koutlemanis, Spiros Paparoulis, and Vassilis Kouroumalis. Tracking persons using a network of rgbd cameras. In *Proceedings of the 7th International Conference on PErvasive Technologies Related to Assistive Environments*, page 63. ACM, 2014.

[25] John Krumm, Steve Harris, Brian Meyers, Barry Brumitt, Michael Hale, and Steve Shafer. Multi-camera multi-person tracking for easyliving. In *Visual Surveillance, 2000. Proceedings. Third IEEE International Workshop on*, pages 3–10. IEEE, 2000.

[26] James Orwell, Simon Massey, Paolo Remagnino, Darrel Greenhill, and Graeme A Jones. A multi-agent framework for visual surveillance. In *Image Analysis and Processing, 1999. Proceedings. International Conference on*, pages 1104–1107. IEEE, 1999.

[27] Anurag Mittal and Larry S Davis. M2tracker: A multi-view approach to segmenting and tracking people in a cluttered scene. *International Journal of Computer Vision*, 51(3):189–203, 2003.

[28] Saad M Khan and Mubarak Shah. A multiview approach to tracking people in crowded scenes using a planar homography constraint. In *Computer Vision–ECCV 2006*, pages 133–146. Springer, 2006.

[29] Francois Fleuret, Jerome Berclaz, Richard Lengagne, and Pascal Fua. Multicamera people tracking with a probabilistic occupancy map. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(2):267–282, 2008.

[30] Jerome Berclaz, Francois Fleuret, Engin Turetken, and Pascal Fua. Multiple object tracking using k-shortest paths optimization. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(9):1806–1819, 2011.

[31] Horesh Ben Shitrit, Jérôme Berclaz, François Fleuret, and Pascal Fua. Multicommodity network flow for tracking multiple people. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(8):1614–1627, 2014.

[32] Rok Mandeljc and Kovacic. Tracking by identification using computer vision and radio.

[33] Bernardin Keni and Stiefelhagen Rainer. Evaluating multiple object tracking performance: the clear mot metrics. *EURASIP Journal on Image and Video Processing*, 2008, 2008.

DVD Content

Contents of the DVD are listed in Table 5.

Table 5: Contents of the DVD

Name	**Description**
thesis	diploma thesis in pdf format.
tracking_system	tracking system folder containing the system, experiments folder with example experiment and data folder with the input data. Installation instructions in README.md
floorball_tracking	visualization of the tracking results on floorball dataset
pets_tracking	visualization of the tracking results on PETS'09 dataset