

FREDRIK NÄSSTRÖM, JONAS ALLVAR,
ERIKA BILOCK, FREDRIK BISSMARCK,
VIKTOR DELESKOG, ROBERT FORSGREN,
HANS HABBERSTAD, FREDRIK HEMSTRÖM,
GUSTAF HENDEBY, JÖRGEN KARLHOLM,
JONAS NORDLÖF, JONAS NYGÅRDS, JOAKIM RYDELL



Fredrik Näsström, Jonas Allvar,
Erika Bilock, Fredrik Bissmarck,
Viktor Deleskog, Robert Forsgren,
Hans Habberstad, Fredrik Hemström,
Gustaf Hendeby, Jörgen Karlholm,
Jonas Nordlöf, Jonas Nygårds, Joakim Rydell

Intelligent spaning 2012-2015

Slutrapport

Titel	Intelligent spaning - Slutrapport
Title	Intelligent reconnaissance – Final report
Rapportnr/Report no	FOI-R--4148--SE
Månad/Month	December
Utgivningsår/Year	2015
Antal sidor/Pages	50 p
ISSN	1650-1942
Kund/Customer	Försvarmakten
Forskningsområde	7. Sensorer och signaturanpassning
FoT-område	Sensorer och signaturanpassning
Projektnr/Project no	E54564
Godkänd av/Approved by	Per Johannesson
Ansvarig avdelning	Sensor och TK-system

Detta verk är skyddat enligt lagen (1960:729) om upphovsrätt till litterära och konstnärliga verk. All form av kopiering, översättning eller bearbetning utan medgivande är förbjuden.

This work is protected under the Act on Copyright in Literary and Artistic Works (SFS 1960:729). Any form of reproduction, translation or modification without permission is prohibited.

Omslagets bildcollage: Översta raden från vänster till höger, den vänstra bilden visar en IR-bild med automatisk måldetektion och den högra bilden visar spaning från stridsfordon 90 (Foto: Janne Bohman/Försvarmakten). Nedersta raden från vänster till höger, den vänstra bilden visar stridsfordon som tar sig fram där minröjarna har röjt väg (Foto: Anne-Lie Sjögren/Försvarmakten) och den högra bilden föreställer TUAV i luften under Mali 02 (Foto: Jimmy Croona/Försvarmakten)

Sammanfattning

I denna rapport presenteras arbetet som utförts i projektet Intelligent spaning. I projektet har forskning bedrivits med syfte att göra framtida sensorsystem mer intelligenta. Genom att vara mer intelligenta kan framtida sensorsystem ge sensoroperatörer bättre situationsuppfattning, högre systemtilltro, mindre risk för felhandlingar (eng. human error) och låg mental arbetsbelastning.

Projektet har bedrivit forskning om algoritmer för automatisk detektion, målföljning, sensorsamverkan, fusion, sensorstyrning och positionering. En effektiv detektionsalgoritm har utvecklats för upptäckt av människor. Den framtagna detektionsalgoritmen är mycket bra på att detektera människor i komplexa miljöer, även stillastående, och har låg sannolikhet för falsklarm. Algoritmen är baserad på maskininläringsteknik, vilket gör att den kan läras att upptäcka även andra objekt såsom fordon, fartyg, helikoptrar etc.

En målföljningsalgoritm har utvecklats för att associera detektioner över tid och skapa målspar. Detta ger möjlighet att estimerar målens position och rörelse även om de inte observeras.

Projektets forskning har visat hur spaning med samverkande styrbara sensorer kan användas för att öka spaningstäckningen och förbättra målsparskvaliteten. Forskningen har också visat hur den egna positionen kan bestämmas genom användning av samverkande sensorsystem. Detta har visat sig vara beräkningsmässigt möjligt i realtid.

Projektets forskning har visat hur informationen från detektions- och målföljningssystemet kan distribueras till ledningssystem, t.ex. SLB (Stridsledningssystem Bataljon), som automatiskt uppdaterar sin lägesbild i realtid. Informationen till SLB distribueras via en NFFI-koppling (NATO Friendly Force Information). Ledningssystemet SWECCIS kan uppdateras på samma sätt. Förmågan att automatiskt uppdatera en lägesbild kan användas av många plattformar inom försvarsmakten såsom fordon, övervakningssystem, helikoptrar, flygplan, UAV etc.

Nyckelord: detektion, målföljning, sensorsamverkan, fusion, sensorstyrning, positionering, maskininläring

Summary

This report presents the work carried out in within the project Intelligent reconnaissance. In the project research has been conducted with the aim to make future sensor systems more intelligent. By being more intelligent, future sensor systems provide sensor operators with better situational awareness, higher system confidence, reduced risk of human errors and low mental workload.

The project has conducted research on algorithms for automatic detection, tracking, collaboration, fusion, sensor management and positioning. An effective detection algorithm has been developed for detection of people. The developed detection algorithm is very good at detecting people in complex environments, and have a low probability of false alarms. The algorithm is based on machine learning technology, which means that it can be taught to detect also other items such as vehicles, vessels, helicopters etc.

A tracking algorithm is developed to associate the detection over time and create target tracks. This gives the opportunity to estimate the targets position and movement, even if they are not observed.

The research project has shown that reconnaissance with cooperating steerable sensors can be used to increase coverage and improve tracking performance. The research has also shown that positioning of the own platform can be determined through the use of collaborative sensors. This has been found to be computationally possible in real time.

The project's research has shown how information from detection and target tracking system can be distributed to a Battle Management Systems (BMS), which automatically updates the situational awareness in real time. The information to the BMS is distributed thru a NFFI coupling (NATO Friendly Force Information). The management system SWECCIS (SWEdish Command and Control Information System) can be updated in the same way. The ability to automatically update a situational picture can be used for many platforms in the Swedish Armed Forces such as vehicles, surveillance systems, helicopters, aircraft, UAVs etc.

Keywords: detection, target tracking, collaboration, fusion, sensor control, positioning, machine learning

Innehåll

1	Inledning	7
1.1	Projektets leveranser	7
1.1.1	Rapporter och FOI Memo	8
1.1.2	Konferensbidrag, journalartiklar och bokkapitel.....	8
2	Intelligent spaning	10
2.1	Spaning och sensorplanering	10
2.1.1	Exempelscenario	10
2.1.2	Metod och design.....	11
2.2	Detektion av människor och fordon	12
2.2.1	Detektering av utbredda och delvis skylda mål	12
2.2.2	Detektering av rörliga mål från en rörlig plattform	16
2.3	Målföljning	23
2.4	Positionering	24
2.4.1	Lokal positionering	25
2.4.2	Samverkande navigering	27
3	Sensorsystem	29
3.1	Krav på sensorsystem med avancerad signalbehandling	29
3.2	Sensorsystemet QWIP.....	29
3.3	Sensorsystemet MSP	30
4	Resultat för signalbehandlingsalgoritmerna	31
4.1	Användning av signalbehandling för förbättrad spaningsförmåga med styrbara EO/IR-sensorer	31
4.1.1	Operatörsstöd med formbaserad måldetektion	31
4.2	Förbättrad spaningsförmåga med samverkande sensorer.....	34
4.2.1	Optimering och värdering av målföljningsprestanda	35
4.3	Presentation av lägesbild med information från signalbehandlingsalgoritmer	38
4.3.1	Användning av geografisk information vid automatisk spaning..	39
5	Fortsatt arbete	40
6	Appendix	41
6.1	Detektering av rörliga mål från en plattform i rörelse	41
6.1.1	Luft till mark.....	41
6.1.2	Mark till mark.....	45
7	Referenser	48

Förkortningar

CLEAR	CLassification of Events, Activities and Relationships
FOV	Field of view
GTSAM	Georgia Tech Smoothing and Mapping
IMU	Inertial Measurement Unit
LIDAR	Light Detection And Ranging
LWIR	Long-wavelength infrared
MHT	Multi-Hypothesis tracker
MSP	MultiSensorPlattform
MSSLab	MultiSensorSimuleringsLab
OSPA	Optimal Subpattern Assignment
QWIP	Quantum Well Infrared Photodetector
ROC	Receiver Operating Characteristic
SIFT	Scale-Invariant Feature Transform
SLAM	Simultaneous Localization And Mapping
UAV	Unmanned Aerial Vehicles

1 Inledning

Syftet med projektet Intelligent spaning har varit att studera hur avancerad signalbehandling kan användas för att göra sensorsystem mer ”intelligenta”. Syftet är att underlätta sensoroperatörens arbete och därmed öka spaningens effektivitet och kvalitet. Målsättningen med de nya signalbehandlingsalgoritmerna som tagits fram i projektet är att de ska bidra till framtida operatörssystem som ska kunna ge bättre situationsuppfattning, högre systemtilltro, mindre risk för felhandlingar (eng. human error) och låg mental arbetsbelastning för operatören.

Utvecklingen inom signal- och bildbehandlingsområdet går snabbt över hela världen. Detta gör att nya automatiska funktioner hela tiden utvecklas. Fyra fördelar med automatiska funktioner är att de vanligen leder till:

- Ökad effektivitet (precision/snabbhet)
- Ökad säkerhet
- Minskad arbetsbelastning (mental och fysisk)
- Förbättrad ekonomi

Även om nya automatiska funktioner utvecklas för signal- och bildbehandling behövs fortfarande en människa som utifrån det övergripande sammanhanget tolkar skeenden och prioriterar sensorresurser. Människan är därför ofta oersättlig precis som i många andra komplexa system. Förr eller senare är det alltid en människa som ansvarar för att styra, justera och omkonfigurera systemet och dess automatiska funktioner beroende på aktuella krav. Systemen måste därför konstrueras så att människan kan använda dem på ett effektivt och säkert sätt.

En vanlig orsak till att operatörer kan ha svårt att hantera komplexa system på avsett sätt är bristen på återkoppling (feedback). Operatören får helt enkelt för lite eller ottydlig information om systemets status och hur den utvecklas [1]. Om systemet och operatören inte kan kommunicera på ett tillfredsställande sätt kan operatören uppfatta att systemet gör helt oväntade saker, s.k. Automation Surprises. Det här var tidigare ett problem inom den civila luftfarten när man började använda automatiserade funktioner som resulterade i en helt ny typ av tillbud och olyckor [2]. Just bristen på återkoppling till operatören var en av orsakerna till de här problemen, vilket gjorde det mycket svårt för operatören att ingripa och göra rätt saker för att hantera störningar. En felinmatad siffra eller ett glömt kommatecken resulterade i att systemet stumt utförde vad det blivit instruerat att göra, hur fel detta än var och utan att visa piloterna resultatet [3].

Det som avgör hur bra ett operatörsgränssnitt fungerar är vilka förutsättningar det ger för att förstå farkostens förmåga att hantera förväntade situationer och att koordinera hur uppgiften ska lösas [4]. Hur omfattande koordinationsproblemet är beror på vilka krav situationen ställer och farkosternas förmåga att upprätthålla en acceptabel prestationsnivå utan operatörens medverkan. Om koordinationsproblemet är svårt så minskar antalet farkoster som operatören kan hantera. Det är därför viktigt med bra gränssnitt för svåra koordinationsproblem så att operatören inte blir överbelastad.

1.1 Projektets leveranser

Projektet har presenterats och demonstrerats vid många tillfällen för Försvarmakten, FMV och industrin. Projektet har även tagit fram flera publikationer, vilka visas i sammanställningen nedan.

1.1.1 Rapporter och FOI Memo

Näsström, F., Deleskog, V., Emilsson, E., Hendeby, G., Karlholm, J. och Rydell J. "Om-världsanalyt inom Intelligent spaning", FOI Memo 4487, 2013.

Näsström, F. "Redovisning av milstolpe kvartal 4 2013 för projekt Intelligent spaning, Försvarmaktens ref.nr. – AF.9220408", FOI Memo 4695, 2013.

Näsström, F., Bissmarck, F., Cedefamn, J., Deleskog, V., Emilsson, E., Forsgren, R., Habberstad, H., Hemström, F., Hendeby, G., Karlholm och J., Rydell J. "Lägesrapport för projektet Intelligent spaning", FOI Memo 4974, 2014.

Näsström, F. "Redovisning av milstolpe kv4 2014 för projekt Intelligent spaning, Försvarmaktens ref.nr. – AF.9220409", FOI Memo 5136, 2014.

Näsström, F. "Från sensor till lägesbild i realtid", FOI Memo 5267, 2015.

Fredrik Näsström, "Intelligent spaning: Slutredovisning i form av seminarium med demonstration", FOI Memo 5499, 2015

1.1.2 Konferensbidrag, journalartiklar och bokkapitel

Saha, S., Hendeby, G. och Gustafsson, F. "Noise Adaptive Particle Filtering: A Hierarchical Perspective", *ISBA Regional Meeting and International Workshop/Conference on Bayesian Theory and Applications (IWCBA)*, 6-10 januari 2013, Varanasi, Uttar Pradesh, India, 2013. (Parallellpublicering: FOI-S--4535--SE)

Deleskog, V. och Hendeby, G. "Intelligent Surveillance with Multiple Sensor Platforms", *Proceedings of Reglermöte*, Linköping, juni 2014. (Parallellpublicering FOI-S--4729--SE)

Roth, M., Hendeby, G. och Gustafsson, F. "EKF/UKF maneuvering target tracking using coordinated turn models with polar/Cartesian velocity", *Proceedings of 17th IEEE International Conference on Information Fusion*, Salamanca, Spanien, juli 2014. (Parallellpublicering: FOI-S--4753--SE)

Saha, S. och Hendeby, G. "Rao-Blackwellized particle filter for Markov modulated non-linear dynamic systems", *Proceedings of IEEE Workshop on Statistical Signal Processing*, Gold Coast, Australien, juli 2014. (Parallellpublicering FOI-S--4758--SE)

Bleser, G., Damen, D., Behera, A., Hendeby, G., Mura, K., Miezal, M., Gee, A., Petersen, N., Maçães, G., Domingues, H., Gorecky, D., Almeida, L., Mayol-Cuevas, W., Calway, A., Cohn, A. G., Hogg, D. C. och Stricker, D. "Cognitive Learning, Monitoring and Assistance of Industrial Work-flows Using Egocentric Sensor Networks", *PLoS ONE*, 10(6), 2015. (Parallellpublicering: FOI-S--5195--SE.)

Bleser, G., Steffen, D., Reiss, A., Weber, M., Hendeby, G. and Fradet, L. "Personalized Physical Activity Monitoring Using Wearable Sensors" i *Smart Health*, Lecture Notes in Computer Science, No. 8700, 99-124, 2015. (Parallellpublicering: FOI-S--5200--SE.)

Reiss, A., Hendeby, G. och Stricker, D. "A novel confidence-based multiclass Boosting algorithm for mobile physical activity monitoring", *Personal and Ubiquitous Computing*, 19(1): 105-121, 2015. (Parallellpublicering: FOI-S--5194--SE.)

Roth, M., Fritsche, C., Hendeby, G. och Gustafsson, F. "The Ensemble Kalman Filter and its Relations to Other Nonlinear Filters", *Proceedings of the 2015 European Signal Processing Conference (EUSIPCO 2015)*, Nice, France, 2015. (Parallellpublicering: FOI-S--5198--SE.)

Saha, S., Hendeby, G. och Gustafsson, F. "Mixture Kalman Filters and Beyond" i *Current Trends in Bayesian Methodology with Applications*, 537-562, 2015. (Parallellpublicering: FOI-S--5199--SE.)

Skoglund, M., Hendeby, G. och Axehill, D. "Extended Kalman Filter Modifications Based on an Optimization View Point", *18th International Conference of Information Fusion*, Washington, DC, USA, 2015. (Parallellpublicering: FOI-S--5197--SE.)

Veibäck, C., Hendeby, G. och Gustafsson, F. "Tracking of Dolphins in a Basin Using a Constrained Motion Model", *Proceedings of the 18th International Conference of Information Fusion*, Washington, DC, USA, 2015. (Parallellpublicering: FOI-S--5197--SE.)

2 Intelligent spaning

Målet med arbetet i projektet har varit att ta fram generisk funktionalitet för samverkande styrbara sensorsystem för att öka effektiviteten och kvalitén på spaningen, men samtidigt underlätta sensoroperatörens arbete. I projektet har vi valt att fokusera på ett spaningsscenario med bemannade stridsfordon och UAV som befinner sig i ett fientligt område med risk för t.ex. eldöverfall.

Forskningen i projektet har inriktats mot styrning av sensorer för optimal informationsinhämtning vid spaning mot markmål (människor och fordon), automatisk detektion av svåra markmål som kan vara partiellt skylda, samt följning och positionsbestämning av multipla mål i komplexa miljöer från rörliga sensorplattformar i samverkan. Geografisk information utnyttjas för förbättrad sensorstyrning, egenpositionering och hotbildsvärdering. Uppgiften kan lösas på många olika sätt, alla med sina för- och nackdelar. Fokus i projektet har varit att visa på fungerande och flexibla lösningar. Huvuddelen av arbetet har varit signalbehandling för stridsfordon, men algoritmerna som beskrivs i rapporten är generella och kan användas även av UAV:er, övervakningssystem, fartyg, helikoptrar, flygplan, robotar etc.

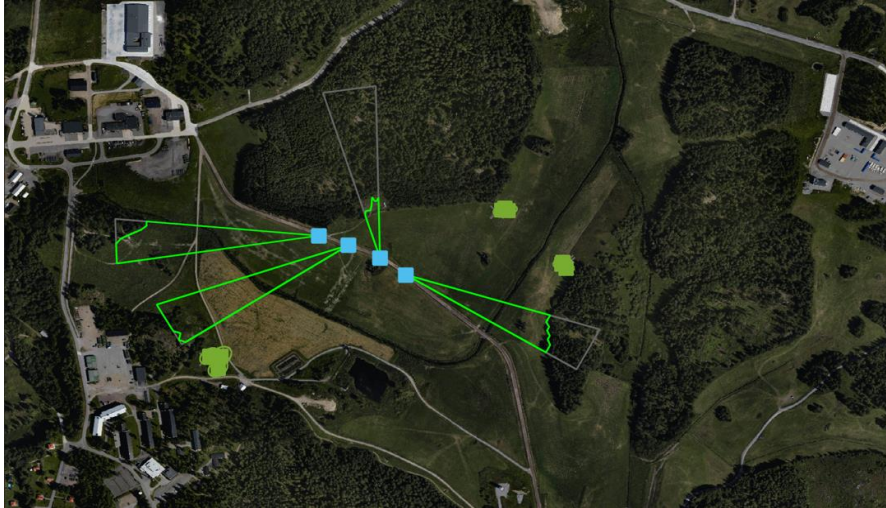
Arbetet i de olika delarna av signalbehandlingskedjan beskrivs i detta kapitel.

2.1 Spaning och sensorplanering

Att bedriva effektiv spaning i miljöer med varierad terräng där flera potentiella hot existerar är ett svårt problem att lösa. En tydlig Lösningsstrategi är därför av hög vikt. Den mest logiska strategin är att dela upp ursprungsproblemet i tydliga delproblem där okända faktorer och förväntad effekt tydligt kan specificeras. Genom att göra denna uppdelning är det möjligt att i slutändan lösa det stora problemet på ett så effektivt sätt som möjligt. Sensorplanering i denna kontext definieras av att de tillgängliga sensorerna kan styras, såsom orientering och zoom-nivå. Sensorplaneringen i sin tur består av en uppsättning av automatiska funktioner, som är utvecklade för att var för sig lösa ett avgränsat delproblem. I huvudsak ska dessa funktioner finnas tillgängliga som verktyg och stöd, vilka tillsammans med sensoroperatörerna kan leverera bättre underlag inför kritiska beslut. Därför måste sensoroperatören sättas i centrum så dess erfarenhet och kunskap kan utnyttjas på bästa sätt. Genom att använda sensorplanering kan tid till upptäckt och eventuell bekämpning av hot minskas speciellt under lång tid då koncentrationsförmåga vanligen blir sämre hos sensoroperatörerna på grund av utmattning och tristess.

2.1.1 Exempelscenario

Ett scenario med eldöverfall har använts för att utvärdera effekten av sensorplaneringen. I scenariot framrycker en spaningsgrupp utmed ett vägnät i ett område med varierande terräng. Spaningsgruppen består av en fordonskolonn utrustade med styrbara spaningssensorer (t.ex. Vapenstation 01). Sensorerna är samverkande i den meningen att sensordata enkelt kan distribueras mellan fordonen, såsom egenposition och detektioner etc. Målet är att upptäcka samtliga okända hot och klassificera dessa som hot, vilka då ska bekämpas, eller neutrala, vilket då inte kräver någon bekämpning. I figur 1 illustreras scenariot där den egna fordonskolonnen framrycker längs vägen i sydöstlig riktning. Vi ser även att tre grupper har upptäckts och klassificerats som neutrala.



Figur 1. Exempelscenario för att illustrera sensorstyrning. En spaningspatrull framrycker utmed vägen samtidigt som sensorerna spanar omkring patrullen. Tre grupper har detekterats och klassificerats som neutrala.

2.1.2 Metod och design

För att kunna utvärdera och utveckla spaningsalgoritmer har ett ramverk för generisk sensorplanering utvecklats under projektets gång. Ramverket har stöd för att styra individuella sensorer burna av en plattform. Plattformen kan antingen vara ett markfordon eller en UAV. För att angripa problemet har två specifika nyckelbeteenden identifierats för att kunna lösa spaningsuppdraget:

- Söka efter mål, sensorn sveper över en fördefinierad vinkelsektor för att hitta nya mål
- Följa mål, sensorn låses på ett specifikt mål för att höja noggrannhet, göra klassificering eller för bekämpning.

Dessa beteenden eller spaningsfunktioner kan då tilldelas sensorer som då blir tillgängliga för sensorplaneringen att utnyttja. Problemet för sensorplaneringen är att avgöra den bästa fördelningen av sensorsystem som ska följa mål respektive söka efter nya mål. Vissa mål är viktiga att hålla uppdaterade och andra tillåts ha sämre positionsnoggrannhet.

För att beräkna den bästa associationen mellan mål och sensorer måste ett mått konstrueras som kan användas som underlag. Underlaget används sedan av en auktionsalgoritm [7] som beräknar de bästa associationerna givet de beräknade måtten. Måttet M_{ij} är definierat som differensen $M_{ij} = R_{ij} - C_{ij}$, mellan den estimerade vinsten R_{ij} och den beräknade kostnaden C_{ij} att associera mål T_i med sensor S_j . De beräknade måtten ställs sedan upp i en associationsmatris, som är input till auktionsalgoritmen, se tabell 1.

Tabell 1. Associations där elementet M_{ij} är det beräknade måttet för att associera mål T_i med sensor S_j .

	S_1	...	S_M
T_1	M_{11}	...	M_{1M}
$\vdots$	$\vdots$	$\ddots$	$\vdots$
T_N	M_{N1}	...	M_{NM}

Vinsten definieras som ökningen av målets positionsnoggrannhet givet målets position samt sensors position och prestanda. Kostnaden i sin tur definieras som hur stor del av ett

helt varv en sensor måste rotera för att ha målet i sitt synfält. Eftersom både målet och plattformen rör sig under tiden måste deras positioner vid den tidpunkten målet beräknas vara i synfältet predikteras så att målet överhuvudtaget kan detekteras av sensorn. Eftersom terrängen är varierande och mål kan vara dolda bakom t.ex. skog har terränginformation tagits med vid beräkning av kostnaden så att sensorer inte associeras med dolda mål. De sensorer som inte associeras med något mål eller inte är konfigurerade att följa mål upprätthåller sin spaning efter nya mål.

Genom att använda denna design kan flera påverkansfaktorer och bivillkor modelleras och integreras i ramverket. Då ramverket till vis del är generellt kan det skräddarsys utifrån specifika spaningssätt och reglementen för att erhålla en så bra spaningseffekt som möjligt.

2.2 Detektion av människor och fordon

De flesta vapensystem inom försvarsmakten har bildalstrande sensorer, vanligen en visuell kamera och en termiska IR-kameror som långvågs-IR (LWIR) eller kortvågs-IR (MWIR).

2.2.1 Detektering av utbredda och delvis skydda mål

På FOI har en måldetektionsalgoritm utvecklats för att automatiskt hitta mål i data från vanligt förekommande EO/IR-sensorer. De detekterade målen kan presenteras i sensoroperatörens skärm som ett stöd eller användas av andra delsystem.

Detektorn är en variant av en Boosting-detektor [5], vilket lämpar sig väl för installation i billig hårdvara. Detektorn delar upp bilden i regioner och scannar dessa för att hitta mål. I de fall där avståndsbedömningen är sämre eller okänd kan detektorn leta efter mål i flera olika skalor.

Till skillnad från många andra detektionsalgoritmer är denna algoritm träningsbaserad och tränas upp på förhand för att detektera specifika mål. Exempel på mål kan vara människor, fartyg, bilar, containrar eller stridsvagnar. Flera detektorer kan köras parallellt, dvs. flera typer av mål går att detektera samtidigt.

I figur 2 visas exempel på detektion av människor i olika storlekar. Människorna längst bort befinner sig på ca 280 meters avstånd och har en höjd av 8-10 pixlar i bilden, människan närmast kameran har en höjd på 30 pixlar. Notera att inga hundar blir detekterade som människor.



Figur 2. Detektion av människor i LWIR, sensorn spanar över Gärdet i Stockholm.

Träning av detektionsalgoritm

Eftersom detektorn är en träningsbaserad detektor behöver man ha träningsdata på den typ av mål man vill detektera. Träningsdata kan exempelvis komma från faktiska spaningsuppdrag eller rena inspelningskampanjer vars syfte är att skapa träningsdata.

Träningsdata är vanligtvis ett markerat område i tidigare inspelat data. För upprättstående människor används markeringar för huvud och fotposition, se figur 3. Mängden träningsdata varierar beroende på mål. Att ta fram markeringar är ofta ett manuellt arbete och för detta har ett markeringsprogram utvecklats i projektet. Attribut som beskriver om personen står upp, hukar sig, ligger ner, cyklar, är skymd etc. kan även sparas och användas vid träningen.



Figur 3. Egenutvecklat annoteringsverktyg för framtagning av träningsdata.

Detektion av upprättstående människor

I det här projektet har fokus varit att detektera i huvudsak upprättstående människor i LWIR. Anledning att vi valt att detektera människor är att insurgenter, skyttesoldater, jägare etc. är mycket vanliga i många konflikter. I projektet har vi byggt upp en databas av inspelat och annoterat material av människor i olika miljöer, väderfall och tidpunkter på året. Totalt har 82 000 bilder på människor tagits fram och används för att träna detektionsalgoritmen.

Detektionsalgoritmen för detektion av utbredda mål har testats på olika sekvenser som samlats in med IR-kamerorna på FOI:s experimentsystem (QWIP och MSP-systemet) för att studera dess prestanda att detektera upprättstående människor. Avstånd för experimentet valdes till: 100, 150, 200 och 250 meter. Inspelningen gjordes med 15 bilder per sekund. Detektorn appliceras på varje bild oberoende av varandra. Ingen rörelseanalys används. Falsklarmsnivåerna var inställda så de undertrycks maximalt, det vill säga inga falsklarm accepteras. Experimentet utfördes i april månad då temperaturen var 8 grader.



Figur 4. Detektion av människa på 250 meters avstånd. Målet är 6-7 pixlar högt.

Följande mått har uppmätts:

- Sannolikhet för detektion per bild
- Sannolikhet för detektion under 1 sekund

Tabell 2. Sannolikhet för detektion för olika avstånd.

Avstånd (meter)	Optik (grader)	Upplösning (pixlar)	Mål (höjd i pixlar)	Sannolikhet för detektion i enstaka bild	Sannolikhet för detektion under 1 sekund
100	20	320 × 240	16	97%	100%
150	20	320 × 240	11	92%	100%
200	20	320 × 240	8	63%	100%
250	20	320 × 240	6-7	37%	73%

Tillförlitliga detektioner uppnås mellan 8-11 pixlar. 10 pixlar är satt för en rimlig gräns för detektion. Ytterligare slutsats för detektionsavstånd för modernare sensorer med högre upplösning återfinns i avsnitt 4.

Falska mål

Detektorn är tränad att se skillnad på mål och bakgrund. I vissa situationer kan varma stenar eller trädstammar ge upphov till falska detektioner. Dessa detektioner kan ibland vara svåra att skilja även för en sensoroperatör, vilket visas i figur 5 (vänster). Den vanligaste källan till falsklarm är för lite träningsdata, den högra bilden i figur 5, visar ett sådant fall när en detektor inte har sett tillräckligt med bilar och tolkar den varma strålkastaren och däckets som en människa. Detta kan motverkas med mer träningsdata.



Figur 5. Exempel på falska mål. Ibland kan varma stenar eller träd se ut som människor, detta kan i sin tur resultera i falska mål (vänster). Varma delar på okända objekt kan orsaka falska mål (höger).

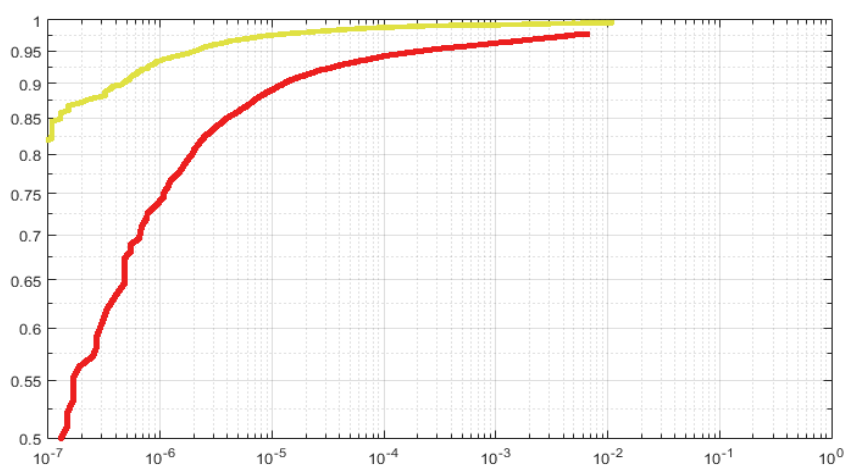
Detektion av delvis skylda människor

Insurgenter som gömmer sig bakom stenar i skog och terräng är av extra vikt att detektera, exempelvis skjutställningar, figur 6. En sätt att detektera delvis skylda människor är att detektera exponerade kroppsdelar, exempelvis överkroppen vid en skjutställning.

Två detektorer har tränats, en för helkroppar (upprättstående) och en med bara data för den övre tredjedelen av kropp (överkropp). I figur 7 visas ROC-kurvor (Receiver operating characteristic) för de båda detektorerna. En ROC-kurva visar förhållandet mellan detektionsprestanda och falsklarm. Överlag är överkroppsdetektorn sämre än helkroppsdetektorn då den ger fler falsklarm. Det är väntat att överkroppsdetektorn är sämre då detektorn bara har en tredjedel av bilddata att arbeta med.



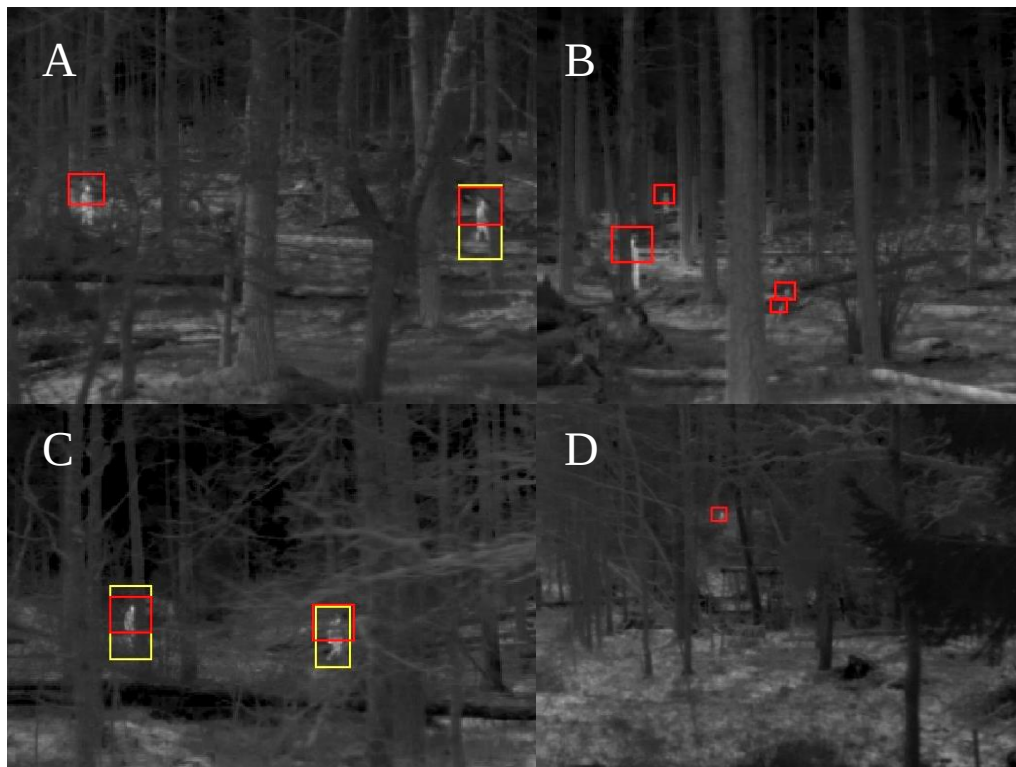
Figur 6. Objekt som skyler sig kan exponera delar av kroppen i exempelvis skjutställning bakom en trädstam.



Figur 7. ROC-kurva för överkroppsdetektor (röd) och helskroppsdetektor (gul).

Även om överkroppsdetektorn överlag har sämre prestanda är tanken att objekt som skyler sig ska kunna detekteras. För att utvärdera detta i praktiken kördes båda detektorerna på en sekvens i skogen där människor är naturligt och delvis skylda av terräng. I figur 8 är båda detektorerna inställda på att acceptera 90% av detektionerna.

En initial observation är att överkroppsdetektorn (röd) hittar fler objekt än helskroppsdetektorn (gul). Dock har överkroppsdetektorn något fler falsklarm än helskroppsdetektorn som i den aktuella sekvensen inte har något falsklarm. Helskroppsdetektorn har även en viss förmåga att detektera när objekten är delvis skylda av grenar och granris, trots att detta inte finns med i träningsdata, figur 8C. I den aktuella sekvensen detekteras ett objekt djupt inne i skogen av överkroppsdetektorn ca 3 sekunder innan motsvarande mål detekteras av helskroppsdetektorn figur 8D. Många falsklarm är orimliga och kan filtreras bort med hjälp av målföljning och användande av riktningssinformation.



Figur 8. I bild A visas bildruta 338 med detektioner från helkroppsdektorn (gul) och överkroppsdektorn (röd). I bild B visas bildruta 321. Överkroppsdektorn hittar fler objekt men har fler falsklarm än helkroppsdektorn (2 objekt, 2 falsklarm). I bild C visas bildruta 226. Helkroppsdektorn kan även detektera objekt skylda av grenar och granris. I bild D visas bildruta 150. Detektion av objekt djupt in i skogen.

2.2.2 Detektering av rörliga mål från en rörlig plattform

En nackdel med formbaserad detektering är att detektionsalgoritmen bara kan känna igen objekt liknande dem den exponerats för under träningsfasen. Objekten måste även täcka ett relativt stort antal bildpunkter för att kunna skiljas från bakgrundsklotter.

Rörelsedetektering är en alternativ ansats där uppgiften är att upptäcka objekt som rör sig relativt den stationära bakgrunden. När kameran är helt stillastående eller endast panorerer (roteras) är detta en jämförelsevis enkel sak. Om kameraplattformen translaterar förflyttas dock även punkter i den stationära bakgrunden på ett komplicerat sätt i bildplanet, vilket medför att en sofistikerad analys krävs för att skilja ut objekt i faktisk rörelse.

Videoföljning

I projektet har en förhållandevis ny metod för följning av punkters rörelse i en bildsekvens studerats. Metoden kallas *partikelpropagering* och gör att betydligt flera punkter kan följas än vad som i allmänhet är möjligt med alternativ metodik, s.k. *särdragsföljning* (eng. *feature tracking*). Detta medför att sannolikheten att upptäcka mål kan förväntas vara större.

Vid särdragsföljning identifieras punkter i bilden, exempelvis hörn, vars lokala omgivning har en sådan struktur att man med hög sannolikhet kan lokalisera dem i efterföljande bilder, även om kameran roterar och translaterar så att orientering och skala ändras. I varje bild söker man efter sådana *intressepunkter* vars lokala omgivningar sedan matchas mot motsvarande uppsättning från föregående bild. När detta upprepas bildas över tiden längre spår som sedan kan analyseras vidare. Den stora fördelen med denna metod är att det är möjligt att matcha vitt skilda vyer. Nackdelen är att antalet stabila intressepunkter med tillräckligt särpräglade omgivningar för att medge säker matchning är relativt lågt i typiska

scener, så att endast några få hundra punkter kan följas. Det är därför stor risk att man missar objekt som inte sticker ut från bakgrunden.

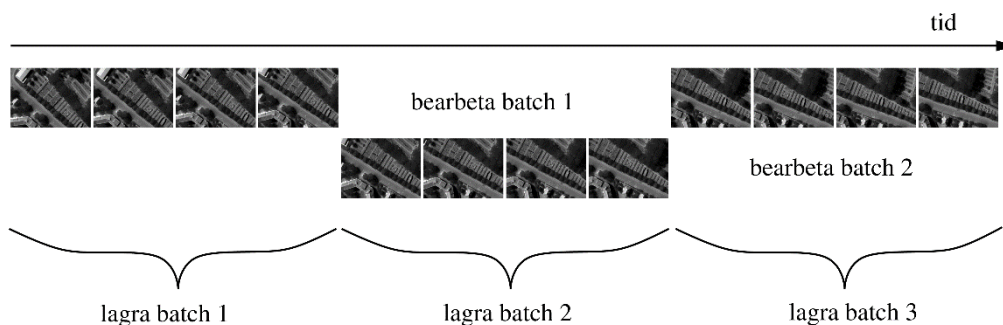
Partikelpropagering bygger på beräkning av *optiskt flöde*, vilket innebär att man skattar samtliga bildpunkters rörelse mellan successiva bilder. I princip görs detta på samma sätt som vid särdragsföljning, men antalet möjliga matchningar minimeras genom att anta att närliggande punkter rör sig på ungefär samma sätt, vilket är sant utom för punkter som ligger nära randen av ett objekt. Punkter vars rörelse skattats inkorrekt identifieras genom att rörelseskattningen görs även i omvänd riktning (bakåt i tiden). Om punkten p_1 i föregående bild matchats mot p_2 i aktuell bild kräver man således även att den bäst matchande omgivningen till denna i föregående bild ligger tillräckligt nära p_1 . På samma sätt som vid särdragsmatchning kan spår genom en bildsekvens skapas genom att länka samman matchningar mellan konsekutiva bilder.

Luft till mark

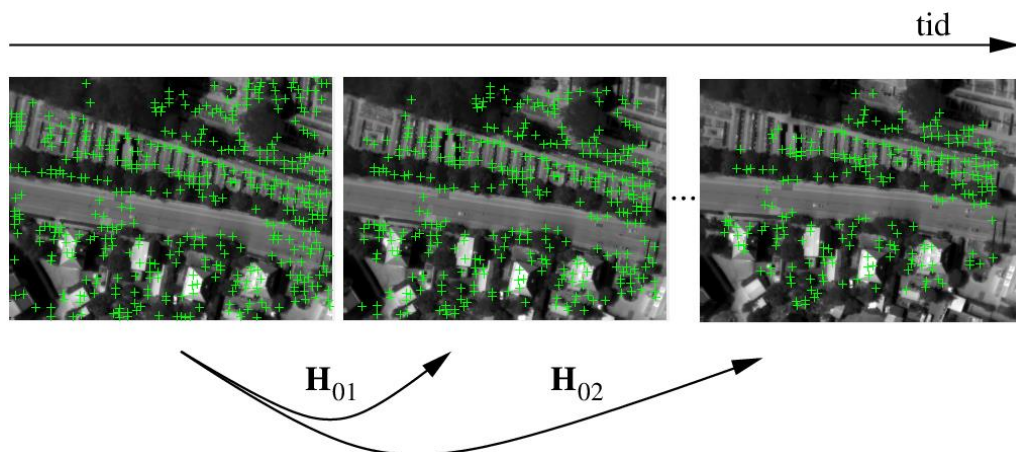
För att detektera rörelse på marken från en flygande plattform (t.ex. en UAV) förslås en ansats (*plan-parallax*) baserad på bildstabilisering. Följande moment ingår:

1. Stabilisera bakgrunden över en delsekvens av överlappande bilder:
 - a. Detektera ett plan i scenen
 - b. Ensa (registrera) successiva bilder mot planet
2. Detektera residualrörelse, dvs. punkter som efter stabiliseringen rör sig relativt planet.
3. Analysera och klassificera residualrörelsen:
 - a. Faktisk rörelse relativt den stationära bakgrunden
 - b. Parallaxrörelse för punkter i bakgrunden utanför det detekterade planet

Bildstabiliseringen sker för successiva delsekvenser, så som visas i figur 9, eftersom det krävs ett signifikant överlapp mellan samtliga bilder. Särdragsföljning används för identifiering av punkter i ett gemensamt plan. I den första bilden i delsekvensen identifieras ett antal intressepunkter (t.ex. hörn), vilka sedan matchas mot motsvarande i nästa bild. Bland alla punktassociationer mellan de båda bilderna behålls den största delmängd som tillsammans rört sig på ett sätt som är konsistent med att de ligger i ett gemensamt plan i scenen. Punkterna följs sedan genom hela delsekvensen, men i varje steg behålls endast de vars rörelse fortfarande är konsistenta med ett plan, se figur 10.



Figur 9. Batchbaserad bearbetning av videosekvens i realtid. I varje delsekvens (batch) stabiliseras (ensas) bilderna mot ett plan synligt i samtliga bilder. Samtidigt som en delsekvens bearbetas lagras nästa i en buffert.



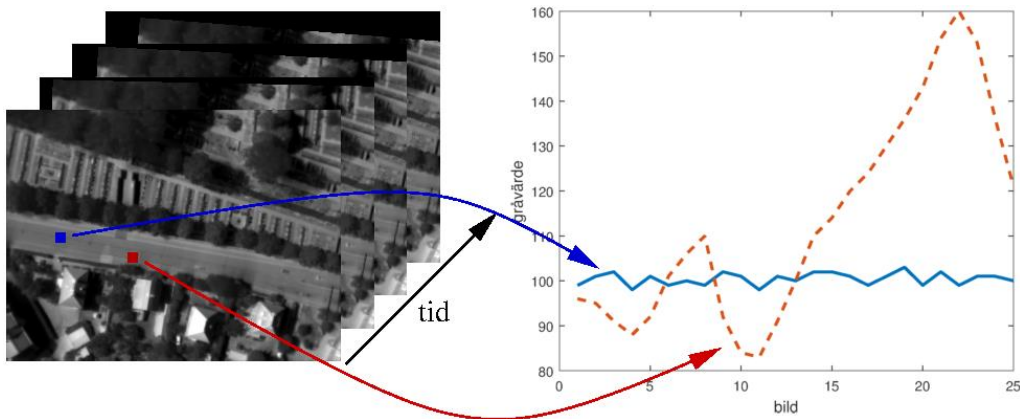
Figur 10. Identifiering av punkter i gemensamt plan. Koordinaterna i bildplanet för punkter vars rörelse är konsistent med att de ligger i ett gemensamt plan i scenen transformeras enligt en avbildning $\mathbf{H}$, en s.k. homografi.

Efter att planet identifierats registreras (ensas) bilderna geometriskt till den första referensbilden, se figur 11. Om bilderna nu läggs ovanpå varandra kommer gråvärdet för punkter i eller nära det skattade planet att variera lite i tiden (blå punkt/kurva i figur 12), medan gråvärdet för övriga punkter i allmänhet kommer att variera betydligt mera (exempelvis röd punkt/kurva, där ett objekt i rörelse passerar). I de punkter där variationen i tiden överstiger ett tröskelvärde startas ett spår, och partikelpropagering används för att följa den lokala omgivningen genom den stabiliserade sekvensen.

För att reducera mängden data grupperas spåren genom klustring. Närliggande spår med liknande rörelse genom sekvensen samlas till ett gemensamt objekt, vars centrumpunkt får representera samtliga spår i gruppen, se figur 13.



Figur 11. Stabilisering av delsekvens. Genom de skattade homografierna kan bilderna ensas (registreras) till den initiala referensbilden.



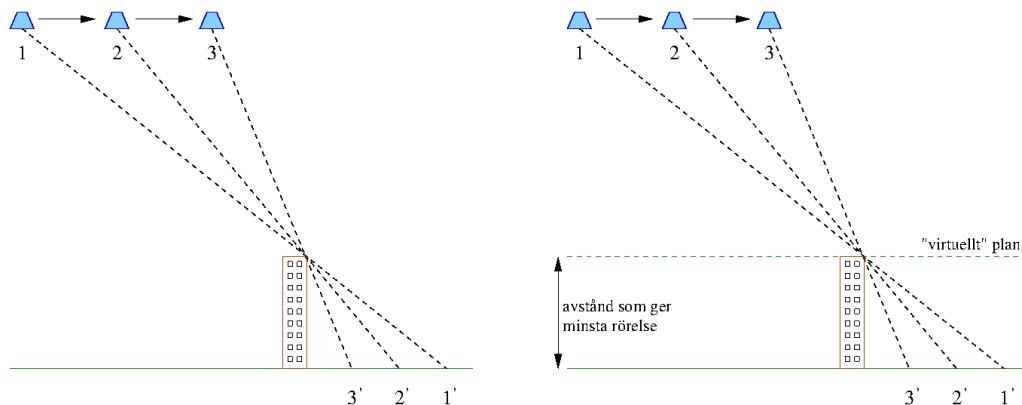
Figur 12. Skillnadsdetektion. Efter stabilisering analyseras variationen i gråvärde över tiden för varje bildpunkt. Punkter som avbildar strukturer i eller nära det skattade planet uppvisar typiskt en liten variation (blå punkt/kurva), medan värdet för övriga punkter varierar betydligt mera (t.ex. den röda punkten/kurvan, som är en bil i rörelse).



Figur 13. Klustring av närliggande spår i gemensam rörelse.



Figur 14. Parallaxrörelse. Sekvensen är stabiliserad mot på markplanet, men det höga tornet har en betydande residualrörelse även efter stabilisering.

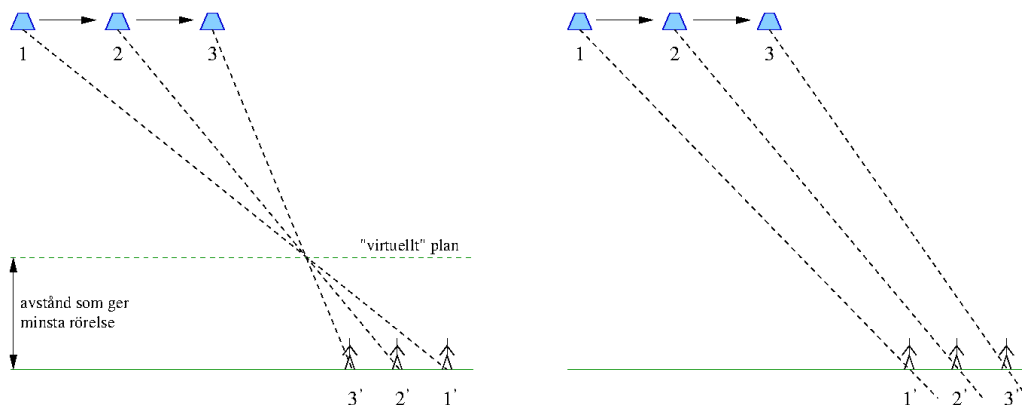


Figur 15. Identifiering av parallaxrörelse. Vänster: När kameran rör sig tycks byggnadens tak röra sig i den mot markplanet stabiliserade sekvensen. Höger: Om bilden istället stabiliseras mot ett plan parallellt med markplanet, men förskjutet till hustakets höjd, är taket fixerat genom sekvensen.

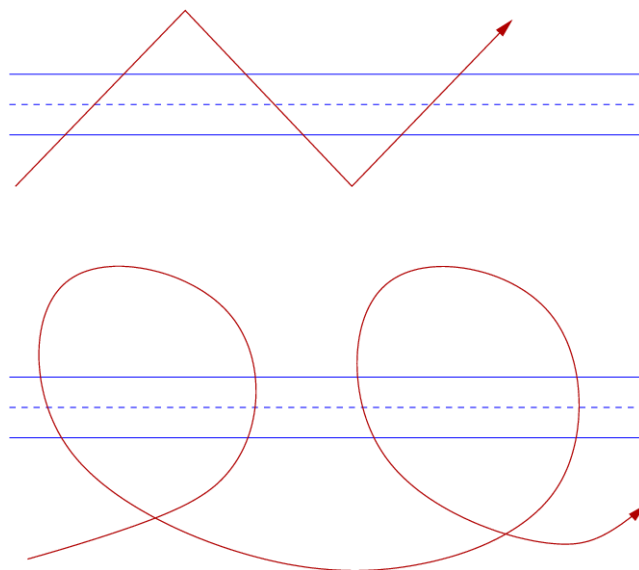
Objekt i faktisk rörelse relativt den stationära markbakgrunden utgör i allmänhet endast en delmängd av dem som identifierats genom ovan beskrivna procedur. Strukturer utanför referensplanet för stabilisering ger upphov till en parallaxrörelse, se figur 14. Situationen illustreras schematiskt i figur 15 (vänster). För att identifiera parallaxrörelse kan man resonera som i figur 15 (höger). Storleken på parallaxrörelsen beror av punktens vinkelräta avstånd från referensplanet. Det betyder att om stabiliseringen istället sker mot ett "virtuellt" plan parallellt med referensplanet, men förskjutet till punktens position, uppstår ingen parallaxrörelse. I projektet har en metod utvecklats för att enligt denna princip identifiera punkter vars rörelse är konsistent med parallax. Metoden beskrivs i Appendix 6.1.1. Några resultat av rörelseklassificeringen visas i figur 16.



Figur 16. Klassificering av residualrörelse som parallax (rött) och verklig rörelse (grönt).



Figur 17. Skenbar parallaxrörelse. Om kamera och mål rör sig i samma eller motsatt riktning och med konstant relativ fart är det omöjligt att skilja objektets rörelse från parallax. I fallet till höger ligger det optimala virtuella planet bortom referensplanet.



Figur 18. Rutter för spaning längs väg.

I figur 17 illustreras att det dessvärre finns viktiga fall där det är principiellt omöjligt att skilja verklig rörelse från parallax, nämligen när kamera och mål rör sig i samma eller motsatt riktning och med konstant relativ fart. Fallet till höger i bilden kanske kan tyckas av akademiskt intresse, men i en starkt kuperad terräng, som exempelvis en stad, kan referensplanet hamna ovanför markplanet, så att det inte alls är orimligt att det optimala virtuella planet ligger bortom referensplanet. För att hantera problemet med skenbar parallax måste sensorplattformens rörelse planeras så att risken minimeras för att kamera och objekt rör sig längs samma riktning. Vid spaning längs en trafikled, t.ex. en väg eller kanal, är rutter liknande dem i figur 18 därför att föredra.

Mark till mark

Den ovan beskrivna metoden att stabilisera delsekvenser med avseende på ett plan i scenen är mindre naturlig när kameran finns på en rörlig plattform på marken. Omgivningen har i detta fall en stor variation i djup och det kan, utom i urban miljö, vara svårt att robust identifiera ett plan i scenen. I denna del av projektet har vi undersökt möjligheten att använda samtidig skattning av scenens 3D-struktur och sensorplattformens rörelse [50] för att detektera objekt i rörelse, se figur 20.

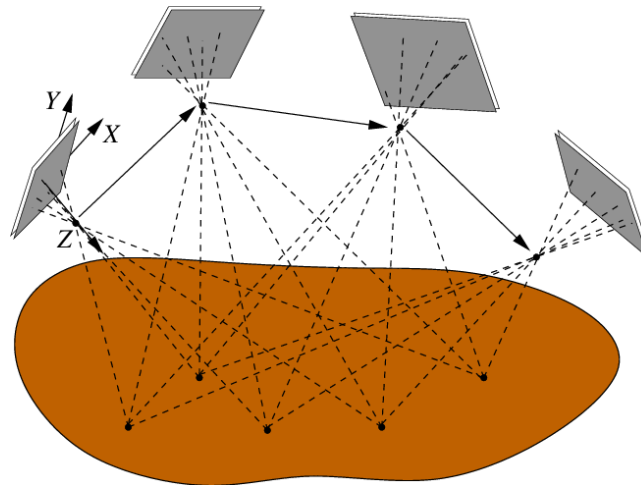


Figur 19. Videoföljning från markfordon. I bilden visas samtliga spår som existerat i minst 5 bilder.

Metoden utgår ifrån följning av så många punkter som möjligt genom videosekvensen, se figur 19. De resulterande spåren används för att bestämma hur sensorplattformen rört sig (roterat och translaterat) relativt en utgångsposition. Denna information kan sedan utnyttjas för att genom triangulering skatta de följda punkternas position i scenen. Eftersom kameran position i varje tidpunkt nu är känd kan de skattade punktpositionerna i rummet

projiceras in i bildplanet och jämföras med de uppmätta spåren. Om skillnaden är stor kan spåret inte höras till den statiska bakgrunden. Metoden beskrivs i detalj i Appendix 0.

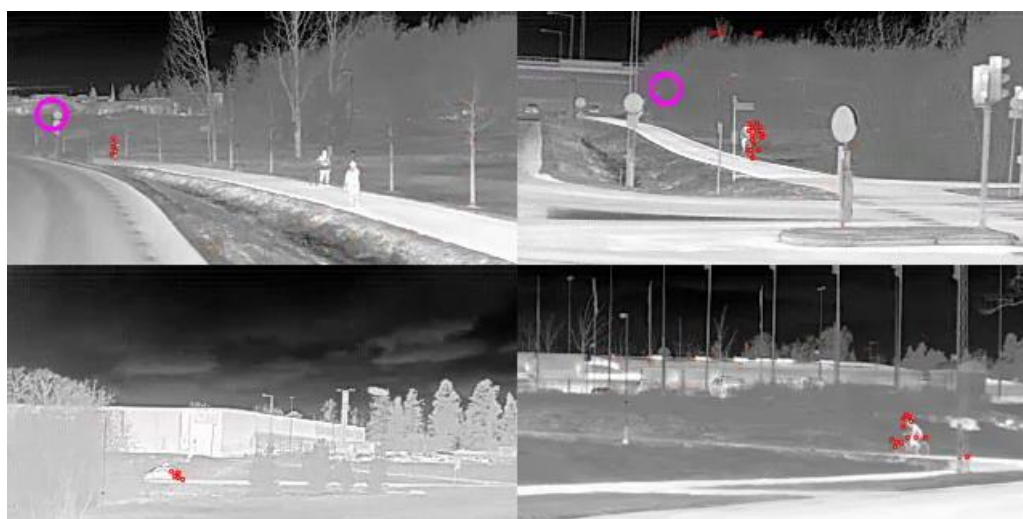
Avstånd kan endast bestämmas som multipler av en referensrörelse som kan väljas som avståndet mellan utgångsläget i första bilden och en godtycklig senare kameraposition. För att bestämma absoluta avstånd krävs information från en annan källa, exempelvis en tröghetssensor eller stereokamera. Detta har dock ingen principiell betydelse för den aktuella metoden eftersom klassificeringen av rörelse är baserad på mätningar i bildplanet.



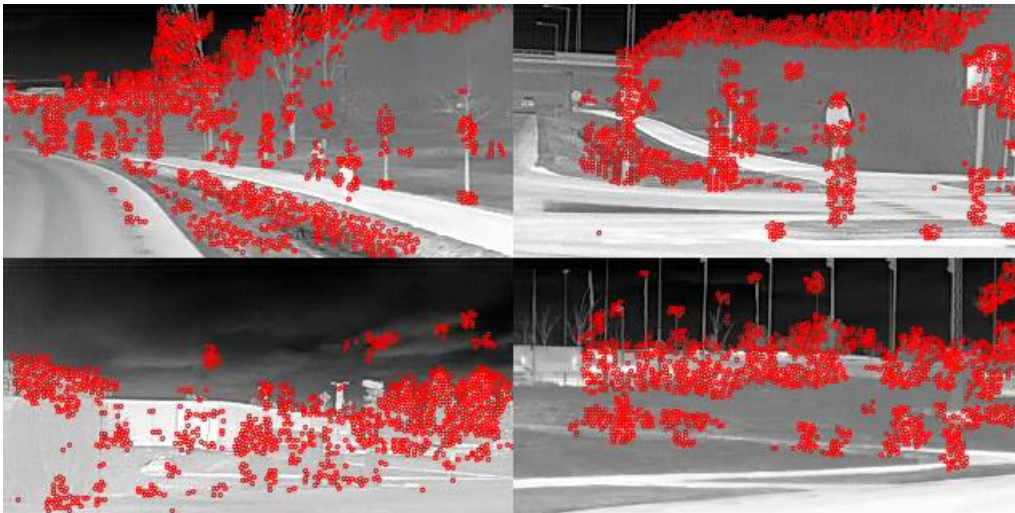
Figur 20. Kamerans rörelse och 3D-positionerna för följda punkter skattas samtidigt genom att minimera det genomsnittliga avståndet mellan punkternas projektioner och de observerade punktkoordinaterna i bildplanen.

Notera också att det, precis som i flygfallet, figur 17, är omöjligt att enbart genom geometrisk analys skilja parallaxrörelse från fallet när kamera och mål rör sig i samma eller motsatt riktning och med konstant relativ fart.

I figur 21 visas några exempel på anomalidetektion, dvs. punkter vars rörelse är inkonsistent med parallax enligt kriteriet ovan. Som referens visas i figur 22 samtliga punkter som följts i minst 5 bilder. Notera i bilden överst till vänster att fotgängarna som förväntat inte detekteras, eftersom deras rörelse är parallell med kamerans. Cyklisten överst till höger detekteras när kameraplattformen bromsar in för att svänga.



Figur 21. Detektering av objekt i rörelse från sensorplattform i rörelse. Ringen markerar projektionen i bildplanet av kameracentrums skattade läge 2 sekunder tidigare (s.k. epipol), vilket ger en uppfattning om kamerans rörelse. I de nedre bilderna panoreras kameran ut från rörelseriktningen, vilket gör att epipolen ligger utanför bilden.



Figur 22. Punktföljning. Samtliga följda punkter vars spår existerat minst 5 bilder (jfr resultat efter parallaxfiltrering som visas i figur 21).

2.3 Målföljning

Med målföljning höjs informationsvärdet i tillgängliga detektioner genom att koppla ihop detektioner över tid och på så sätt skapa målspar som beskriver hur ett objekt har rör sig över tid (främst position och hastighet). Genom att ta med tidsaspekten får man ökad möjlighet att undertrycka falska detektioner genom att de inte beter sig som mål. De producerade spåren kan sedan användas som insignal till mer avancerade metoder. I skenet av detta är en bra målföljare en vital komponent i många sensorsystem.

Med denna bakgrund har inom projektet genomförts forskning inom målföljningsområdet, både på systemnivå och på för systemet viktiga delkomponenter. Arbetet, vilket beskrivs mer detaljerat nedan, har resulterat i ny och förbättrad funktionalitet i den egenutvecklade MATLAB-toolboxen ”Multi-Target Matlab Tracker” samt bättre förståelse för dess olika komponenter.

Multi-Target Matlab Tracker är ett allmänt ramverk för multimålföljning implementerat i MATLAB. Toolboxen bygger på en state-of-the-art multihypotes-tracker (MHT [6, 7]) och har stöd för alla moderna följefilterstyper:

- *kalmanfilter* (KF [8]) och *extended kalmanfilter* (EKF [9]);
- *unscented kalmanfilter* (UKF [10, 11])
- *informationsfilter* (IF) och *extended informationsfilter* (EIF); samt
- *partikelfilter* (PF [12, 13]).

Under den gångna perioden har dessutom ramverket utökats med *probability hypothesis density* (PHD) filter [14] kapacitet genom samarbete med annan parallell verksamhet.

Den verksamhet som bedrivits inom målföljningsområdet sammanfattas i stora drag med följande punkter:

- Multi-Target Matlab Tracker har konsoliderats, robustifierats och snabbats upp.
- Toolboxen har anpassats för enkel användning som en komponent i MSP och för att förse arbetet med sensorplanering lämplig indata.
- Inledande studier kring automatisk utvärdering av målföljningslösningar.

- Studier av egenskaper och förbättrade metoder för följefilter och andra relaterade frågeställningar.

Totalt har verksamheten inom projektet Intelligent Spaning resulterat i 2 journalartiklar, 2 bokkapitel och 7 konferensbidrag inom målföljningsområdet.

De två första punkterna ovan rör båda arbete med målföljningstoolboxen med syfte att göra den mer användbar i demonstrationer och som bas för mer avancerade förmågor. Resultatet har, förutom att ha använts i talrika demonstrationer, även dokumenterats och presenterats som ett konferensbidrag [15].

En frågeställning som ofta återkommer är hur man värderar multipel målföljning, dvs. vilken tracker som genererade bäst spår. Det är inte en trivial fråga, eftersom spårning ger upphov till olika typer av fel – noggrannhet, kardinalfel, ofullständigheter och förväxlingar. Man måste tillföra en subjektiv värdering beroende på tillämpningen för att kunna vikt de olika felen mot varandra. Vi har, parallellt med annan verksamhet, tagit fram ett verktyg för att automatiskt kunna utvärdera och jämföra målföljningsprestanda. Det ger också möjlighet att titta på olika mått som förekommer i litteraturen.

Utöver detta har det i mindre omfattning studerats hur olika problemstrukturer kan användas för att på ett mer effektivt sätt implementera t.ex. de följefilter som ligger till grunden för moderna multimålföljningsalgoritmer. Ett resultat är att genom att dela upp problemet i olika delar, med olika egenskaper, där varje del kan lösas med speciellt anpassade metoder och sedan kombinera lösningarna på ett korrekt sätt är det möjligt att lösa problem som annars hade varit för komplexa och tidskrävande för att lösa alls. Arbetet beskrivs i en serie av två konferensbidrag [16, 17] och har även resulterat ett bokkapitel [18] som ger en sammanställning av liknande angreppssätt.

I konferensartikeln [19] jämförs olika rörelsemodeller och följefilter (EKF och UKF varianter) med varandra för ett antal benchmarkbanor. Resultatet visar att problemet är förhållandevis förlåtande, och andra överväganden i de flesta fall ska vara avgörande för det slutgiltiga designvalet. De flesta rörelsemodeller för mål tar inte hänsyn till att mål ibland rör sig i begränsade områden trots att sådan information kan vara mycket informativ. Därför studeras ett sätt för hur man kan angripa sådana problem i konferensbidraget [20]. Den resulterande målföljningsalgoritmen visar på god prestanda i fall då klassiska lösningar har svårt att leverera resultat. Den studie som presenteras i konferensbidraget [21] tittar mer ingående på teorin för det vanliga EKF och knyter samman den med optimeringslära. På så sätt skapas nya möjligheter att lösa problem med olika begränsningar, t.ex. vad gäller var objekt kan befinna sig, som inte faller inom det ramverk som EKF fungerar för.

Framöver förespås möjligheten att hantera stora problemställningar få en växande betydelse. Det kan t.ex. handla om att följa ett stort antal mindre mål som till stor del påverkas av samma externa störningar. I dag finns begränsad möjlighet att hantera den typen av situationer, som även för dagens datorer blir för komplexa för att kunna utföras. Arbetet i konferensbidraget [22] syftar till utnyttja *ensemble Kalman filter* (EnKF), som tidigare används t.ex. för prediktera väder och havsströmmar, för att hitta mer effektiva lösningar med lägre beräkningskomplexitet.

Slutligen har målföljningstekniker och klassificeringsresultat tillämpats för att skatta den fysiska statusen hos personer, och för att hjälpa och till att stötta dem i komplexa tekniska moment och undvika arbetsskador samtidigt som den totala arbetsbelastningen utvärderas. Arbetet finns dokumenterat i två journalartiklar [23, 24] och ett bokkapitel [25].

2.4 Positionering

Att veta position och pekriktning för sina sensorer är viktigt för de flesta avancerade sensorssystem, både civila och militära. Detta gäller för både positionering och inriktning av spaningssensorer på t.ex. stridsfordon, vilket behövs för inriktning av indirekt eld och för

att kunna skapa en gemensam lägesbild mellan flera sensorbärande enheter. Ett annat exempel är positionering av egen personal, vilket är en viktig del av blue force tracking och stridsledning. Idag dominerar GPS-teknik nästan helt för positionering; det är dock viktigt att komma ihåg att GPS är störkänslig, fungerar sämre i bergig och urban terräng, och att en GPS inte ensam kan bestämma riktning.

Hur avancerade navigationsutrustningen som kan användas avgörs ofta av det sensorbärande systemet. Till följd av begränsat utrymme och vikt har små plattformar ofta mindre avancerad utrustning. GPS kan under vissa förutsättningar ge en tillräcklig bra positioneringsnoggrannhet, men kan inte ses som ett tillförlitligt system i svåra terrängförhållanden (så som småbruten terräng eller tät skog), i urban miljö, eller om precisionskraven är mycket höga. I dessa situationer behöver GPS-navigeringssystem ofta stötas eller kompletteras för att erhålla nödvändig precision och tillförlitlighet.

Ett alternativ till GPS navigation är *simultaneous localization and mapping* (SLAM, [26, 27]). SLAM är en typ av algoritmer som försöker positionera sig samtidigt som de kartlägger sin omgivning. Ett SLAM-system använder andra primära sensorer (t.ex. kameror, tröghetsensorer och/eller LIDAR) och kan ge god lokal positionering. Stötts systemet av GPS eller inmätning av globala landmärken kan positionering i globala kartkoordinater uppnås.

Projektet Intelligent Spaning har därför under den gångna treårsperioden arbetat med positioneringsmetoder med fokus på att kunna positionera stridsfordon med god global positionsnoggrannhet. Det första steget för att uppnå detta har varit att bygga vidare på metoder som ger god relativ noggrannhet som sen kan låsas upp mot kända punkter i terrängen för att på så sätt åstadkomma global positionering. Problemet har delats upp i två delar vilka beskrivs mer nedan:

1. Lokal positionering; att ge ett lokalt konsistent resultat i realtid.
2. Samverkande positionering; att ge ett globalt korrekt resultat, givet den lokala lösningen från en eller flera sensorbärare och t.ex. kända landmärken eller GPS mätningar.

Arbetet har resulterat i:

- En prototyp av en sensorplattform för lokal positionering baserad på ett stereokamerapar och tröghetssensorer, stöttat av GPS när tillgänglig för att ge global positionering.
- En kortare studie av hur *ultra wideband* (UWB) radio kan stötta ett lokalt positioneringssystem.
- En första utvärdering av en lovande metod för samverkande global positionering.

De resultat som producerats under projektiden har resulterat i ett prisbelönt konferensbidrag [31].

2.4.1 Lokal positionering

I projektet har ett positioneringssystem ursprungligen avsett för inomhusbruk anpassats för positionering av fordon. Det ursprungliga positioneringssystemet bestod av visuella stereokameror och tröghetssensorer [28, 29, 30]. Positioneringssystemet har modifierats genom att avståndet mellan kamerorna i stereoparet ökats, vilket förbättrar möjligheten att bestämma avståndet till punkter på längre avstånd från sensorn. Det nuvarande systemet visas i figur 23.



Figur 23. Illustration av det nya SLAM-systemet anpassat för utomhusbruk. Här monterat på ett vridbord tillsammans med en visuell och en IR-kamera.

Det nuvarande systemet består av följande sensorer:

- två Point Grey Blackfly kameror monterade i ett stereopar;
- en Xsense MTi-10 IMU; och
- en Ublox LEA-6T GPS-mottagare.

Positioneringssystemet är baserat på EKF-SLAM-metodik, vilket innebär att SLAM genomförs med ett extended Kalman-filter (EKF). Filtret gör det möjligt att väga samman IMU-data (tredimensionell acceleration och vinkelhastighet) med bildinformation från kamerorna i stereoparet. Filtret beräknar systemets relativa position, hastighet, orientering och IMU-enhetens drift. För att öka noggrannheten i positionsskattning identifierar systemet landmärken i varje bildpar från kamerorna. I denna kontext är landmärken punkter i bilderna som är väldefinierade, identifierbara och stationära. För att hitta lämpliga punkter som uppfyller dessa krav har en SIFT-algoritm (Scale-Invariant Feature Transform) använts. Genom att följa dessa landmärken i ett antal bildsekvenser kan systemets skattning av IMU-data förbättras, vilket i slutändan gör positioneringssystemet mer robust och kapabelt att leverera högre precision i positionsskattningen över en längre tidsperiod.

För att öka systemets precision ytterligare styrs kamerorna av en synkroniseringspuls från IMU-enheten. Detta möjliggör för systemet att veta den exakta tiden då bilderna togs och vilka IMU-signaler som påverkas av den uppdaterade felskattningen.

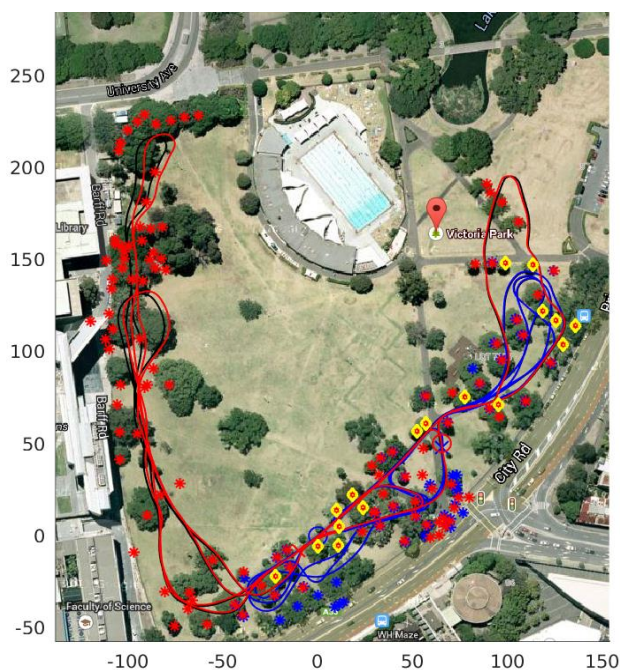
Genom att integrera ett satellitpositioneringssystem, så som GPS, kan systemet relatera sin relativa position till ett globalt koordinatsystem. Fördelen med det integrerade systemet är att, till skillnad från system som enbart förlitar sig på satellitpositionering, kan systemet fungera med hög noggrannhet även i områden utan satellittäckning eller i situationer då satellitsignalerna har blivit störda. Ytterligare en fördel med integrationen är att systemet kan tidsstämpla alla signaler och positionsskattningar med satelittid som har en mycket hög precision. Sådan tidsstämpling kan användas av flera oberoende enheter utan någon intern tidssynkronisering, vilket gör att systemet kan tillämpas i flera användningsområden och av system i samverkan.

Inledande tester för att bedöma systemets nuvarande precision och prestanda har gjorts och kommer fortgå under systemets vidareutveckling. I ett fortsatt arbete planeras att kombinera/ersätta systemets stereokamera med t.ex. en skannande laster (LIDAR).

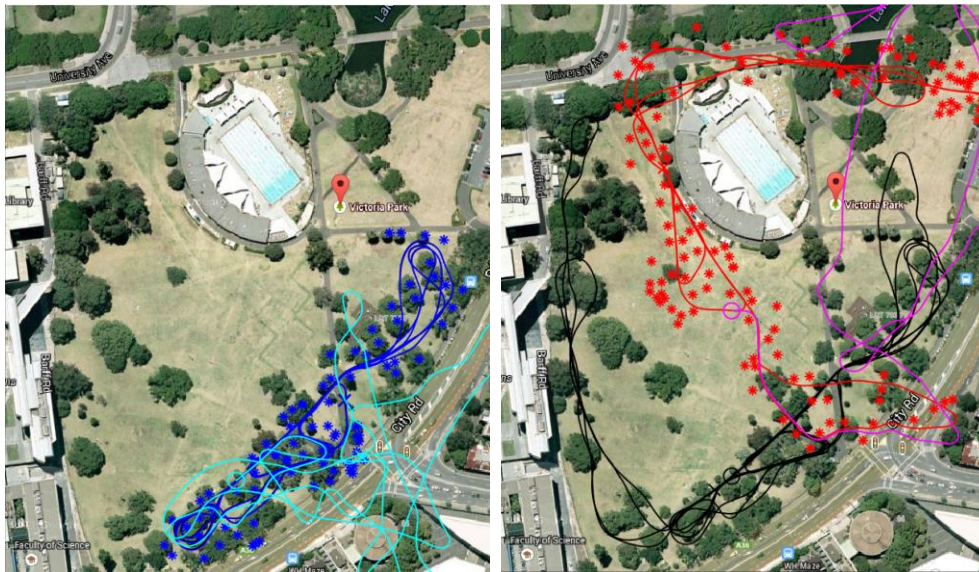
Vidare har begränsat arbete gjort med monokular-SLAM (SLAM baserat på en vanlig kamera utan stereomätningar) stöttat med avståndsmätningar från UWB radiosändare och -mottagare. Det resulterande konferensbidraget nominerades till "Best Paper Award" på konferensen där det presenterades [31]. UWB ses som en intressant metod för att stötta SLAM-lösningar på mindre skala.

2.4.2 Samverkande navigering

För att arbeta med sensorsamverkan integrerat i det nya mätsystemet krävs att sensorerna arbetar i ett gemensamt koordinatsystem. För de fall när GPS inte är tillgänglig samt för att förbättra skattningen av sensorernas orientering har vi studerat användning av gemensamma landmärken i projektet. En öppet tillgänglig kod, GTSAM, från Georgia Tech USA [32] har använts vid analysen. Initiala tester på benchmarking data från Victoria Park ger gott resultat (se figur 24). Datat från Victoria Park är ett väletablerat SLAM-benchmark och har samlats in av Eduardo Nebot et al. vid Australian Centre for Field Robotics, University of Sydney [33]. Victoria Park benchmark är baserat på data från en laser-scanner där trädstammar detekterats. I projektet pågår arbete för att istället använda stereodata från det egenutvecklade positioneringssystemet beskrivet ovan. Idén bygger på att utnyttja GTSAMs grafstruktur för att enkelt men statistiskt noggrunda korrekt knyta ihop navigeringslösningar från flera plattformar. I testet på data från Victoria Park har vi skapat ett virtuellt andra fordon genom att klippa en längre sekvens från ett enskilt fordon i två delar (se figur 25). Algoritmen som är bas för navigeringen ”Incremental smoothing and mapping” (iSAM, [34]) bygger på att samtidigt optimera hela trajektorian och landmärkenas positioner. Optimeringen visar sig ha en gles matrisstruktur vars lösning i GTSAM effektivt kan representeras och optimeras i en grafstruktur som iterativt ombalanseras under lösningens gång. Lösningen i iSAM erhålls i nära realtid, vilket är tillräckligt eftersom fordons enskilda positioneringssystem ger realtidslösningar och uppgiften enbart består i att kontinuerligt ensa de två koordinatsystemen.



Figur 24. Gemensam navigering av två (virtuella) fordon, fordon2 och dess landmärken i rött har lokaliserats mot fordon 1 baserat på identifierade gemensamma landmärken. Gula romber representerar detekterade träd identifierade i båda fordonens kartor, den blå linjen och markörer representerar fordon1. Den svarta linjen är en referenslösning baserat på hela det ursprungliga datasetet, fordon1 och 2 är virtuella i den meningen att de är tagna från samma fordon vid två olika tidpunkter i datasetet. (Kartdata ©2015 Google)



Figur 25. Vänster: Trajektorier och detekterade träd i blått för det virtuella fordon1 inpassad mot google maps. I bilden illustreras också fordonsoodometrin som en cyanblå linje. Höger: Trajektorier och träd rött representerat i lokala koordinater för det virtuella fordon 2. I bilden illustreras också totallösningen över det verkliga fordonets hela trajektorier i svart samt fordon2s odometri i lila. (Kartdata ©2015 Google)

Resultaten på benchmarkdata är lovande och visar på stora möjligheter för framtiden. Framförallt visar det att samverkan och kartdelning över relativt långa perioder är beräkningsmässigt möjligt i realtid. Inledande försök med datainsamling med det lokala positioneringssystemet i kapitel 2.4.1 har gjorts och kommer i det fortsatta arbetet utvärderas på samma sätt som mot det laserbaserade benchmarking-datat från Victoria Park. Målet är att ha ett integrerat system där lokal positionering agerar "front-end" och levererar stabila landmärken till ett system för samverkande positionering i realtid och över relativt stora områden.

3 Sensorsystem

3.1 Krav på sensorsystem med avancerad signalbehandling

I försvarsmakten finns många styrbara EO/IR-sensorer som t.ex. Vapenstation 01, UAV Korpen och UAV Örnen. Ett problem med de flesta sensorsystem i Försvarsmakten är att de är slutna system, vilket gör det svårt att få ut data som inte har komprimerats. Sensorsystemen i FM har ofta analoga gränssnitt som PAL eller NTSC-format för att hämta ut sensordata, men dessa är i praktiken endast lämpade för visning på bildskärm.

Många IR-sensorer kan mäta med en temperaturupplösning på 14 bitar per pixel. Dessa rådata brukar ofta komprimeras till 8 bitar om de ska presenteras på en bildskärm. För att detektionsalgoritmerna ska få så bra prestanda som möjligt, t.ex. tidig målupptäckt och färre falsklarm, är det bäst om sensordata komprimeras så lite som möjligt. Det underlättar även för detektionsalgoritmer om IR-sensorn är radiometrisk eftersom man i en sådan försöker att mäta in alla strålkällor och kompensera och kalibrera för detta, vilket gör att man kan mäta temperaturen i scenen.

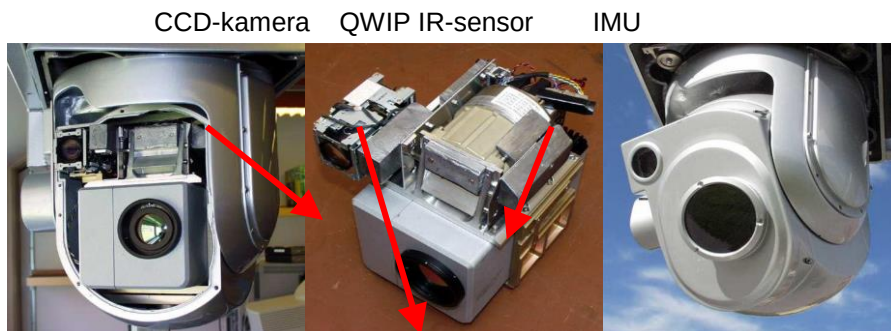
För signalbehandling som t.ex. målföljning är det mycket viktigt att veta vid vilken tidpunkt som detektioner kommer ifrån och vilken position och orientering som sensorerna har vid denna tidpunkt. Av denna anledning behöver de olika sensorerna synkroniseras.

Av dessa anledningar har vi i projektet vidareutvecklat egna sensorsystem istället för att använda EO/IR-sensorer från FM. Sensorerna i våra mätsystem är dock valda för att efterlikna sensorer som används i många militära sensorplattformar.

3.2 Sensorsystemet QWIP

Arbetet med det styrbara sensorsystemet QWIP började 2001 och systemet har kontinuerligt använts och vidareutvecklats inom projektet. QWIP-systemet består av en IR- och visuell sensor i en styrbar gimbal. IR-sensorn i QWIP-systemet är en ThermaCAM SC3000 från FLIR Systems, vilket är en långvågs infraröd sensor med en QWIP (Quantum Well Infrared Photodetector) fokalplansarray. Detektorn i sensorn består av 320×240 pixlar och dess känslighetsområde är ungefär $8.0\text{--}9.2\ \mu\text{m}$. Den monterade optiken ger ett synfält på $20^\circ \times 15^\circ$.

I QWIP-systemet finns även en CCD-kamera som är en vanlig analog enchips PAL video-sensor med ett synfält på mellan 2.7° - 48° . Kamerorna är monterade i en gyrostabiliserad kameragimbal som är styrbar i pan/tilt-led. Utöver kamerorna finns en tröghetsensor (IMU) och vinkelgivare för pan/tilt samt en optisk vinkelmätare. Sensorsystemet QWIP visas i figur 26.



Figur 26. Sensorsystemet QWIP. T.v.: Gimbalen med avmonterad front. Mitten: Det inre sensor-paketet, med IR-sensor, videosensor och IMU. T.h.: Gimbalen med monterad front.

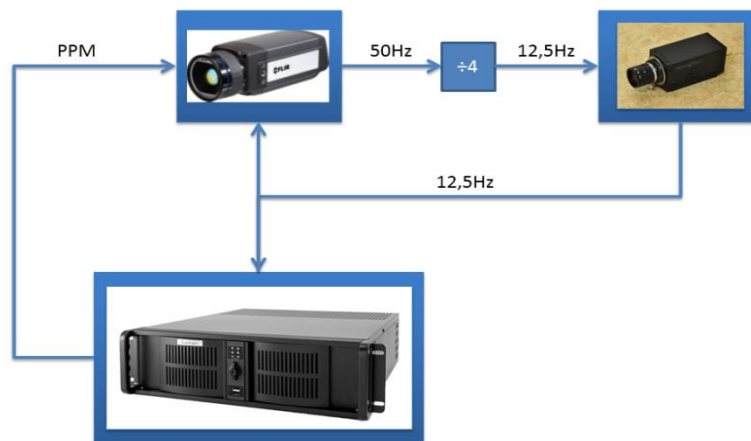
3.3 Sensorsystemet MSP

Sensorsystemet på QWIP-systemet börjar nu bli gammalt och arbete har därför påbörjats med en ersättare till QWIP-systemet som kallas MSP (Multisensorplattform). Det nya sensorsystemet består av en värmekamera FLIR 655sc, en visuell kamera PointGrey ZBR2 bestyckad med Sony ICX687 CCD, 1/1.8" och Tamron M118FM16 optik för FOV: $25.9^\circ \times 19.5^\circ$ som är monterad på ett pan/tilt-vridbord, FLIR PTU-D300E ISM, se figur 27. En stor fördel med vridbord är att det är mycket lättare att byta ut sensorer, vilket gör MSP-systemet mer flexibelt. Sensorerna är valda för att efterlikna sensorer som används i många militära sensorplattformar som t.ex. vapenstationer och UAV:er.



Figur 27. Styrbara sensorsystemet MSP med värmekamera FLIR 655sc, visuell kamera PointGrey ZBR2 monterade på Pan/tilt-vridbord, PTU-D300E ISM från FLIR Motion Control Systems.

För signalbehandling som t.ex. målföljning är det mycket viktigt att veta vid vilken tidpunkt som detektioner kommer ifrån och vilken position och orientering som sensorerna har vid denna tidpunkt. Arbete har därför utförts för att synkronisera sensorerna i sensorsystemet. För närvarande har synkroniseringen av kamerorna skett genom att använda IR-kamerans 50 Hz synkroniseringspuls för utgående synkronisering. Den har sedan delats ner med en faktor fyra till en önskvärd bildfrekvens på 12,5 Hz i systemet. Den visuella kameran konfigureras så att den genererar en synkpuls för varje tagen bild, vilken används för att markera motsvarande bild i IR-kameran. För att kunna relatera bilderna till en tidpunkt, dvs. tidkodning, har insamlingsdatorm försetts med hårdvara för registrering och distribuering av GPS-tid där en PPM-signal (puls per minut) genereras och refereras i IR-datasekvensen genom markering i motsvarande bild, se figur 28.



Figur 28. Principschema för synkronisering av videobilder.

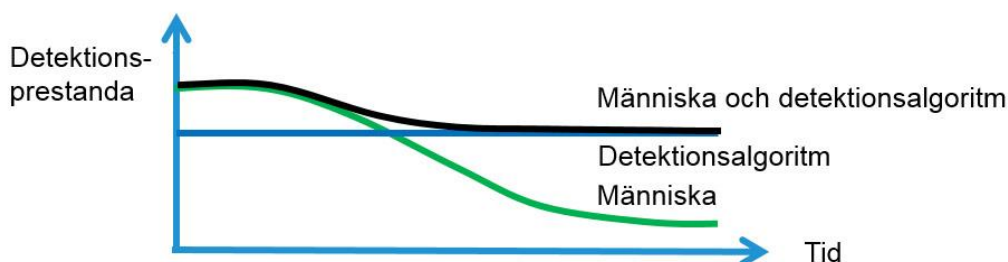
4 Resultat för signalbehandlingsalgoritmerna

I projektet har olika signalbehandlingsalgoritmer utvecklats för att förbättra spaningsförmågan hos både flygande och fordonsburna plattformar med styrbara EO/IR-sensorer. I detta avsnitt redovisas hur dessa signalbehandlingsalgoritmer kan användas för att förbättra spaningsförmågan mot markmål.

4.1 Användning av signalbehandling för förbättrad spaningsförmåga med styrbara EO/IR-sensorer

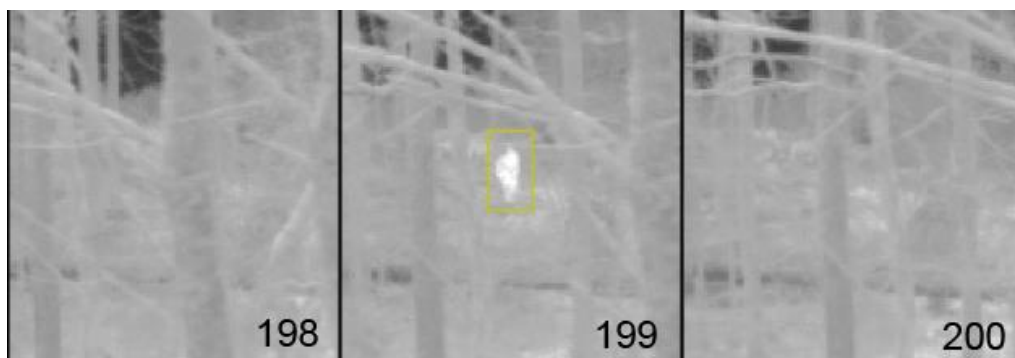
4.1.1 Operatörsstöd med formbaserad måldetektion

Sensoroperatörer är generellt mycket duktiga att hitta mål i terrängen. Uppmärksamheten för människor avtar dock med tiden. Hur snabbt uppmärksamheten sjunker för en människa beror på system och situation. Studien [35] visar att uppmärksamheten sjunker inom 15 minuter, men om arbetsuppgiften är krävande redan inom fem minuter, studie [36]. Figur 29 visar en illustration av att en sensoroperatör är bättre än ett system i ett initialskede, men över tiden avtar uppmärksamheten och systemet är bättre. Allra bäst spaningsprestanda bör således erhållas om sensoroperatören samverkar med ett automatiskt måldetektionssystem.



Figur 29. Illustration av upptäcksprestanda över tid mellan en maskin och sensoroperatör.

En automatisk detektionsalgoritm kan vara ett stöd för en sensoroperatör för att bibehålla eller förbättra prestanda. Ett exempel visas i figur 30, där en sensoroperatör, under färd, spanar in i skogen. Människan som ses i bildruta nummer 199 är bara synlig i ett enda ögonblick, i föregående och nästa bildruta är människan skyddad bakom ett träd. En operatör med uppmärksamheten på en annan plats i bilden kan lätt missa en sådan situation, och på så sätt bli notifierad att det möjligtvis kan finnas fler objekt.



Figur 30. I vissa situationer är mål bara synliga i några ögonblick. I denna situation döljs målet av träd. En automatisk detektionsalgoritm ser varje bild och kan stödja sensoroperatören.

I vissa fall kan en automatisk måldetektionsalgoritm hitta objekt tidigare än en operatör. I ett försök som visas i figur 31 visas ett exempel där detektionsalgoritmen upptäcker människor i terrängen 3-4 sekunder före en människa som tittar på sensordata, vilket är ett bra exempel på hur signalbehandling kan förbättra spaningsförmågan.



Figur 31. Detektion från styrbar IR-sensor på ett fordon i rörelse.

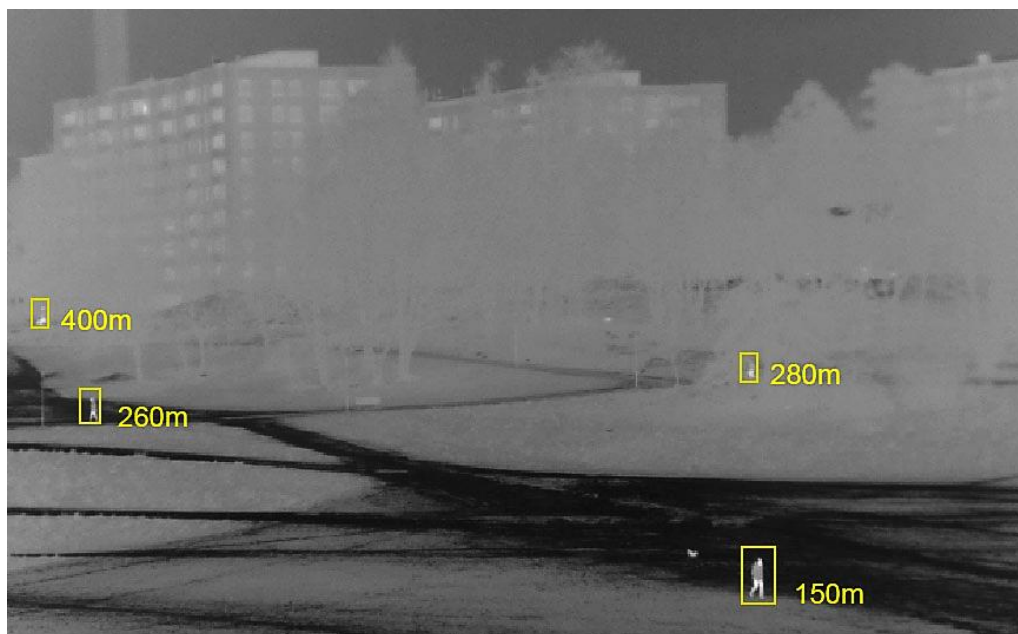
Upptäcktsavstånd

Ett viktigt prestandamått vid upptäckt av mål är upptäcktsavståndet. För upprättstående människor får man enligt avsnitt 2.2, pålitliga detektioner ner till 10 pixlar. Under bra förhållanden är det även möjligt att detektera människor med 6-7 pixlar på målet. Vilket avstånd det mot svarar beror på sensorupplösningen och förstoringen på aktuell optik. Andra faktorer som spelar in är dämpningen i atmosfären, exempelvis blir det kortare upptäcktsavstånd vid dimma. Nedan följer en lista på teoretiska upptäcktsavstånd för några vanliga sensor konfigurationer. Många värmekameror går att köra i flera zoomnivåer kallade ”wide” respektive ”narrow”. Värt att notera är området för detektioner minskar med att synfältet minskar.

Tabell 3. Avstånd för automatisk måldetektion för olika IR-kameror.

Sensor Resolution	Horizontal Field of View	Target Size (pixels)	Distance (meter)	Type
640 × 480	15 (wide)	10 px	436 m	LWIR
640 × 480	5 (narrow)	10 px	1300 m	LWIR
640 × 480	13 (wide)	10 px	490 m	LWIR
320 × 240	10 (wide)	10 px	329 m	LWIR

Figur 32 visar ett exempel från en okyld LWIR-kamera (FLIR 655sc), där en människa detekteras ca 400 m bort med en optik på 15 grader.



Figur 32. Detektion av människor på ett avstånd av ca 150-400 m med okylld värmekamera i urban miljö (Gärdet i Stockholm).

Rörelsedetektion

Rörelsedetektion är ett viktigt komplement till formbaserad detektering. Till skillnad mot formbaserad detektering kan metoden upptäcka alla typer av objekt i rörelse, inte bara sådana som liknar dem den exponerats för under en träningsfas. Även kamouflerade och signaturanpassade objekt, som kan vara mycket svåra att skilja från bakgrunden i stillbilder, kan upptäckas. I princip kan mål som täcker väsentligt färre bildpunkter upptäckas än vad som är möjligt med en formbaserad detektor mot en klotterbakgrund.

När kameran är helt stillastående eller endast panorerar (roteras) kan etablerad robust teknik från *videoövervakning* (adaptiv bakgrundsmodellering) användas för detektering av rörelse. Denna metodik kan vara av stort värde för att avlasta personal vid tillfällig halt eller vid fast spaning från observationsplatser och får betecknas som jämförelsevis mogen.

Om kameraplattformen är i rörelse (translation) förflyttas även punkter i den stationära bakgrunden på ett komplicerat sätt i bildplanet (parallaxrörelse), vilket medför att en mer sofistikerad analys krävs för att skilja ut objekt i faktisk rörelse. Vid spaning från luft mot mark (eventuellt även mark-till-mark i urban miljö) kan styckvis bildstabilisering mot en plan yta i scenen användas, följt av filtrering av återstående parallaxrörelse (*plan+parallax*-ansats). Det finns flera sätt att göra denna filtrering, varav en metod beskrivits i denna rapport. En framgångsrik implementering kräver en noggrann skattning av referensplanet för stabilisering (exempelvis med användning av *bundle adjustment*), samt en punktföljningsmetod som kan generera många långa spår utan associationsfel och drift.

Mark-till-markspaning från en sensorplattform i rörelse är generellt betydligt svårare än luft-till-mark, eftersom man här i regel har en mycket större djupvariation i scenen, med ett stort spann av hastigheter som inte enkelt kan kompenseras för genom bildstabilisering. Vår uppfattning är att det generella fallet kräver mera forskning för att uppnå en större robusthet. Förmodligen kan utnyttjande av information från en tröghetssensor fäst vid kameran och användning av ett stereopar av kameror för kompletterande avståndsskattning bidra till förbättrade prestanda. Det är dock värt att notera att ett viktigt specialfall kan hanteras på ett robust sätt liknande luft-till-markfallet, nämligen när spaning sker mot ett avlägset terrängavsnitt med ett smalt synfält (flera hundra meter med ett 5-gradersobjektiv [37]). I detta fall kan bakgrundens rörelse modelleras med ett plan och mål i rörelse identifieras som punkter som inte passar in i den modellen.

4.2 Förbättrad spaningsförmåga med samverkande sensorer

Genom att använda flera sensorer kan spaningen förbättras i ett område eller så kan spaningsområdet utökas med bibehållen spaningstäckning. Med flera sensorer ökar dock svårigheten för en beslutsfattare att tolka all sensorinformation. Samverkan med flera sensorsystem som automatiskt kan följa flera mål kan därmed vara mycket användbart för att ge en förbättrad lägesbild. Givet att sensorerna samverkar kan viktig information distribueras mellan operatörer i olika fordon, vilket möjliggör att en gemensam lägesbild delas mellan samtliga fordon. Detta bör kunna bidra till en mer effektiv spaning och därför har samverkansförmåga varit en viktig aspekt som har studerats inom projektet.



Figur 33. Vänster: Gemensam lägesbild där målspar och de egna spaningsfordonen visualiseras på en karta. Höger: Målspar och detektioner har överlagrats på sensordata.

Vi illustrerar här ett exempel på effekten av samverkande sensorsystem. Scenariot som använts handlar om en spaningssituation där två stillastående fordon ska spana mot ett öppet fält. I spaningsområdet rör sig två okända människor. Fordonen är utrustade med varsin styrbar IR-sensor, motsvarande den i Vapenstation 01. Sensorn har förmåga att detektera människor, vilket utnyttjas av en multimålföljare för att skapa målspar. Då detektioner från samtliga sensorer finns tillgängliga för multimålföljaren kan en uppsättning målspar skapas där detektionsinformationen utnyttjas på ett sådant sätt att målsparsprecisionen kan maximeras med avseende på osäkerheter i detektioner och målrörelse. Spårn distribueras sedan till varje fordon där de kan visualiseras i en gemensam lägesbild, se vänster bild i figur 33, samt överlagras på sensordata, vilket visas i den högra bilden.



Figur 34. Vänster: Målspar med positionsosäkerhet när ett av målen har observerats av en sensor. Höger: Samma målspar med positionsosäkerhet efter det att målet även har observerats av den andra sensorn.

En tydlig effekt av att sensorerna samverkar erhålls när båda sensorerna detekterar samma mål. Eftersom en ensam sensor har dålig uppfattning om målpositionen i djupled, se vänster bild i figur 34 kommer de måldetektioner från den andra sensorn att fusioneras så att den nya informationen av målet utnyttjas för att öka precisionen i djupled genom ”kryssspejling”, se högra bilden.

4.2.1 Optimering och värdering av målföljningsprestanda

Med optimering av målföljningen kan vi få givna sensorer att samverka för att uppnå bästa möjliga resultat. Man kan t.ex. optimera parameteruppsättningar eller beteenden (rörelsemönster) hos sensorer och sensorplattformar. En förutsättning är att det finns ett prestandamått att optimera för, ett mått som speglar det kvalitativa resultatet.

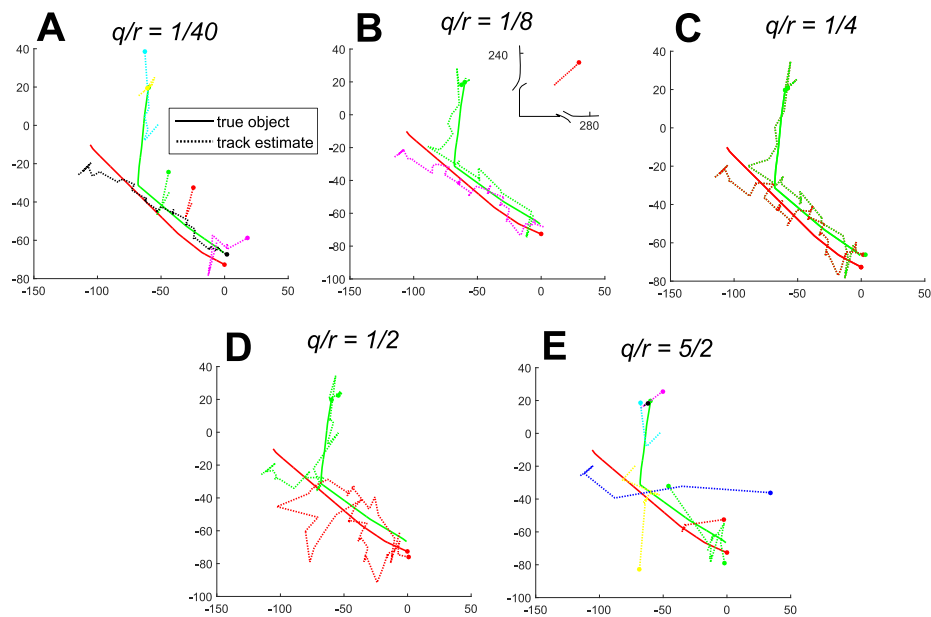
Målföljning kan ge upphov till flera olika typer av fel, när detektionerna är för få eller när antaganden om sensorer och objekt inte stämmer med verkligheten. Spår kan ha dålig noggrannhet, uppkomma av falska detektioner, förväxlas med ett annat, vara ofullständigt, eller saknas för ett objekt i spaningsbilden. Detta gör det svårt att jämföra prestanda hos olika målföljare, eftersom det är som att jämföra äpplen och päron: de presterar ofta olika på olika typer av fel. Man måste då först göra en relativ värdering av dessa fel, beroende på vilken tillämpning man har.

Ett antal mått förekommer för utvärdering av prestanda för multipel målföljning, men det är fortfarande ett aktivt forskningsområde. Ett verktyg för att kunna mäta multipel målföljningsprestanda har vidareutvecklats [38]. Med detta verktyg kan vi jämföra prestanda mellan olika målföljare, optimera parametersättningar, och jämföra lämpligheten hos olika mått.

Nedan följer två tillämpade exempel på optimering av målföljning, först en parameteroptimering, och sedan en beteendeeoptimering.

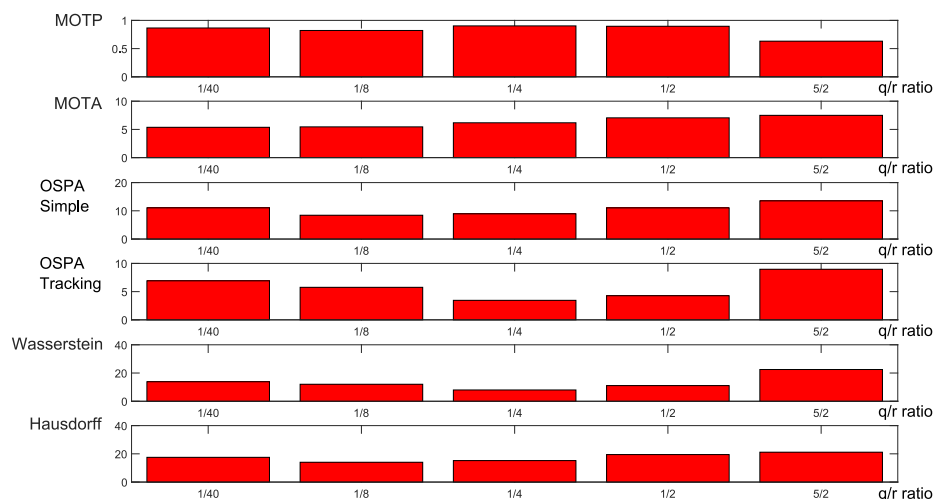
Exempel 1 - parameteroptimering

Här visar vi hur utvärderingsverktyget kan användas för att utvärdera prestanda hos en målföljare i ett scenario snarlikt det beskrivet i 4.2. För enkelhets och tydlighets skull använder vi bara en sensor, men resultatet och tillämpningen är analogt med ett scenario med flera sensorer. I det valda scenariot panorerer en IR-kamera ut över ett fält, med en aktiv persondetektor (motsvarande figur 33). Två personer rör sig mot och passerar varandra, en av dem byter sedan riktning. Vi vill generera spår utifrån detektionerna, och vi vill veta vilken parameter vi ska sätta på kvoten mellan osäkerheterna i rörelse- (q) och observations (r)-modellerna, för att vår tracker ska ge så rimliga tolkningar som möjligt. q^2 och r^2 är variansen i både x (longitud) och y (latitud) för respektive modell. Kvoten q/r avgör hur tungt förutsägelsen av objektets rörelse ska väga gentemot nya observationer.



Figur 35. Parametersättningens effekt på målföljningen. 5 olika parametersättningar A-E, med ökande av kvoten q/r (mer vikt för rörelsemodellen), ger olika spårtolkningar. De sanna objektens rörelse är visade med heldragna linjer (grönt och rött), med punkter i slutet av trajektorierna. Spår (tracks) visas med streckade linjer med unik färg för varje spår. I fall C så svarar spåren mot de sanna objekten - två personer som möts. I D sker en förväxling av spåren där personerna möts.

Målföljningsalgoritmen är ett Extended Kalman Filter (EKF) med en kamera-observationsmodell (perspektivkamera) som kördes offline på sensordata med 5 olika värden på kvoten q/r . Vi kan se mycket olika resultat beroende på parametersättning. Vid för låg och hög kvot klarar inte målföljaren av att associera detektionerna, och därför genereras många spår från samma objekt (figur 35A/E). Ett optimum kan vi se i figur 35C, där spåren stämmer överens med de sanna objekten. Något lägre tal (figur 35B) ger nästan rätt resultat, medan med något högre (figur 35D) så förväxlas spåren.



Figur 36. Kumulativt fel (normaliserat med antal tidssteg) för fem olika q/r -kvoter (svarar mot figur 35), för sex olika prestandamått. (Samtliga mått har utvärderats med en cutoff distance på 20 m).

Finns det ett kvantitativt mått som speglar de kvalitativa resultat vi kan utvärdera i figur 35? Vi mätte upp det kumulativa felet i scenariot för 6 olika felmått: CLEAR MOTP & MOTA [39], Optimal Subpattern Assignment (OSPA) "Simple" (per tidssteg [40], OSPA

”Tracking” [41], Wasserstein- [40] och Hausdorff-avstånd [40]. I figur 36 ser vi att OSPA Tracking och Wasserstein-avstånd stämmer mest väl överens med vår kvalitativa bedömning. Där får vi lägst fel för de ”bästa” spåren ($q/r = 1/4$, figur 35C) och högst för de med ”sämsta” ($q/r = 1/40$ respektive $5/2$, figur 35A/E).

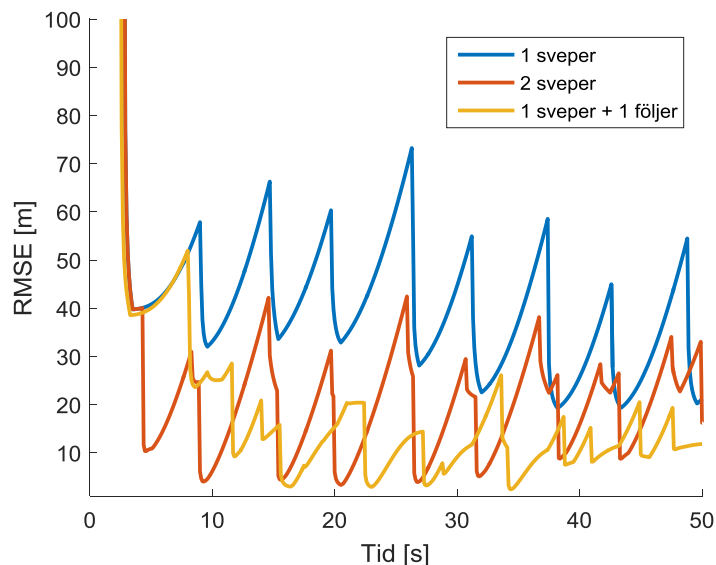
Exemplet här är bara en illustration – i en skarp tillämpning studerar man många fler parameteruppsättningar, och utsätter målföljaren för mycket mer sensordata. Vi får dock redan här en fingervisning om att metoden är tillämpbar. Mera detaljerade studier av felmått och hur de ska tillämpas och jämföras pågår.

Exempel 2 - beteendeoptimering

Det faktum att sensorerna är styrbara ger möjligheten att styra dessa på ett sådant sätt att inhämtning av ny målinformation optimeras med avseende på ökad målsparprecision. I ovan beskrivna exempel kommer tre olika styrstrategier att studeras. Den första strategin använder endast en sensor och låter den svepa över hela spaningsområdet. Den andra strategin utnyttjar att två sensorer kan samverka samtidigt som de sveper över varsin vinkelsektor i spaningsområdet. Den tredje strategin bygger på en frigående strategi där dessutom en sensor tillåts att låsa på mål och följa det kontinuerligt över tid. På så sätt kan målpprecision bibehållas istället för att som i de två föregående strategierna minskas när inte någon sensor observerar målet. Här används det inom projektet utvecklade sensorstyrningsramverket för att avgöra vilka styrkommandon som ska skickas till vardera sensor för att maximera informationsinhämtning.

För att utvärdera de olika styrstrategierna har tre simuleringar genomförts, en för vardera strategi. I de följefilter som beräknar målens estimerade tillståndsvektor (position och hastighet) beräknas samtidigt den estimerade osäkerheten i tillståndsvektorn. Denna osäkerhet motsvarar därmed den estimerade precisionen i målposition. I figur 37 illustreras hur det skattade medelfelet i målposition, för ett gemensamt observerat mål i de tre fallen, varierar över tiden vilket kan tolkas som den förväntade målprecisionen. Måttet som används är roten av det kvadratiska medelvärdesfelet även kallat *Root Mean Square Error* (RMSE). RMSE visar spridningen i avstånd, mätt i meter, från målets korrekta position.

Vi ser att den strategi som ger bäst resultat är den när en sensor förutom att svepa över sin vinkelsektor kan låsa på ett mål. Positionsfelet hålls därmed på lägre nivåer än för när de andra två strategierna appliceras. Effekten av samverkan är stor när man jämför de två första strategierna med varandra. Den högsta uppnådda precisionen med en sensor är omkring 30 m, men för två sensorer kan den uppnå uppemot 3 m, vilket motsvarar tio gångers förbättring. Att precisionen varierar beror på att sensorerna sveper eller byter mål att låsa på, vilket gör att precisionen avtar under tiden då målet inte observeras av någon sensor.



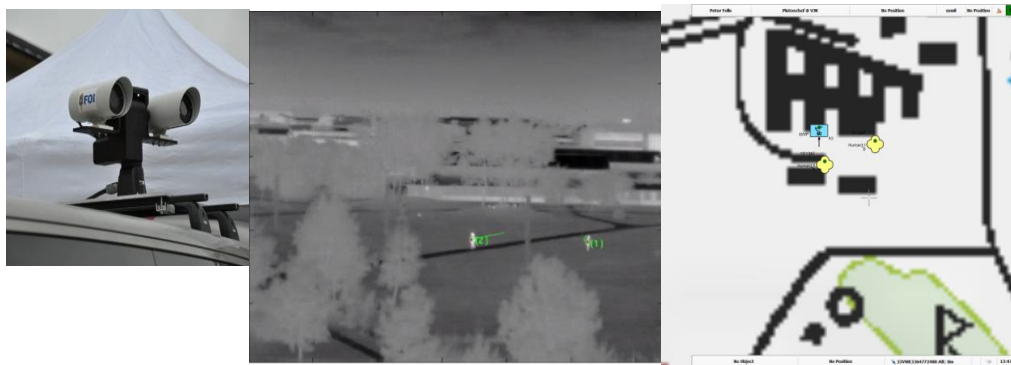
Figur 37. Det estimerade medelfelet RMSE över tid för varje styrstrategi mätt i meter. Den strategi som ger lägst fel är den tredje där en sensor tillåts följa ett specifikt mål.

4.3 Presentation av lägesbild med information från signalbehandlingsalgoritmer

Manöverkrigföring är en central del i Försvarsmaktens syn på stridens genomförande och nyttjande av sina väpnade resurser [42]. OODA-loopen (Observe, Orient, Decide, and Act) eller "Boys cykel" ligger till grund för manöverkrigföringen [43]. Syftet med manöverkrigföringen är att ta sig genom beslutscykeln, OODA-loopen, snabbare än fienden. Förmågan att kunna upptäcka, bedöma, fatta beslut och agera i högt tempo och med god precision (rätt beslut) är därför mycket viktigt.

I projektet har en koppling skapats mellan sensorsystemet MSP och ledningssystemet SLB (Stridsledningssystem bataljon) för att möjliggöra uppdatering av ledningssystemet SLB med målinformation från en IR-sensor i realtid. Detta bör vara mycket intressant för Försvarsmakten på grund av att man på detta sätt kan ta sig genom OODA-loopen betydligt snabbare.

Målspar genererade från de automatiska detektions- och målföljningsalgoritmerna översattes till informationsmodellen som används i MSSLab [44]. Denna information togs sedan in i scenario- och spelverktyget NetScene för att slutligen föras över till SLB, vilket visas i figur 38. Att denna koppling inte gjordes direkt till SLB från sensorsystemet beror på att det redan fanns en koppling från NetScene till SLB och att man med denna lösning även kan leverera simulerad information från MSSLab till SLB. Kopplingen mellan NetScene och SLB skedde via en NFFI-koppling (NATO Friendly Force Information).



Figur 38. Vänster: ett sensorsystem med bl.a. en IR-kamera. Mitten: IR-bild med två personer som automatiskt har detekterats och följs med en målföljningsalgoritm. Höger: SLB (Stridsledningssystem bataljon) med sensorns position och de detekterade personerna markerade.

4.3.1 Användning av geografisk information vid automatisk spaning

Allt mer geografisk information som satellitdata, vägnätsinformation etc. finns nu tillgängligt och denna information är mycket viktigt för att erhålla bra resultat vid spaning. Eftersom man med en vanlig EO/IR-kamera inte får någon avståndsinformation är geografisk information viktigt.

Vid målföljning behöver spår initieras och då är det bra att veta hur långt bort målet är relativt kameran. I terräng med t.ex. böljande landskap och varierande terräng är det svårt att initiera målsåret korrekt relativt kameran, vilket kan ses till vänster i figur 34. Med en laser kan avståndet mätas upp noggrant, men samtidigt finns en risk att motståndaren upptäcker lasern med en laservarnare.

I situationer när flera sensorer finns tillgängliga kan positionsbestämningen förbättras genom triangulering som till höger i figur 34. För att förkorta tiden när målsåret har dålig precision kan en höjdkarta användas. Höjdkartan innehåller information som underlättar för målföljningsalgoritmen att bättre skatta en initial målposition. Inom Sverige har vi väldigt bra höjdkartor, vilket självklart ska utnyttjas av dessa algoritmer i den grad det går.

Ur ett sensorstyrningsperspektiv är all slags information om omgivningen intressant för att kunna generera så optimala styrstrategier som möjligt med avseende på den tillgängliga informationen så som målsår, sensorer, terränginformation och höjddata. Terränginformation kan användas för veta vilken terrängtyp det är dvs. skog, öppen slätt etc. Sådan information tar vi människor lätt till oss genom att observera omgivningen och kan planera därefter. För att en sensorstyrningsalgoritm ska generera styrstrategier som tar hänsyn till den omgivande terrängen kan sådan data vara till stor nytta. I det sensorplaneringsramverk som har börjat utvecklas inom projektet har enkel terränginformation tagits i beaktande. Den terrängsinformation som behandlas är om sikten skymms av träd, hus och berg, se figur 1. Informationen utnyttjades av planeringsalgoritmen så att endast sensorer som kan observera ett mål eller inom kort framtid kommer observera målet antingen för att fordonet rör sig eller om målet rör sig associeras att följa målet. Detta medför att sensorerna allokeras på ett mer resurseffektivt sätt.

5 Fortsatt arbete

De framtagna signalbehandlingsalgoritmerna har redan idag mycket bra funktionalitet, men utvecklingen inom området är mycket snabb.

Fokus på arbetet har varit signalbehandling för stridsfordon, men algoritmerna som beskrivs i rapporten är generella och kan användas även användas av UAV:er, övervakningssystem, fartyg, helikoptrar, flygplan etc. Vissa av algoritmerna behöver dock tränas om för att kunna användas i andra scenarier.

Inom sensorplanering har lovande resultat framkommit under projektets gång. På konceptuell nivå krävs mer arbete för att få olika plattformar att effektivt samarbeta mot ett gemensamt mål. Det kommer dessutom framöver vara viktigt att omsätta de lovande resultaten till praktiskt tillämpliga lösningar baserade på resultat från resterande komponenter i signalbehandlingskedjan.

Arbete med automatisk måldetektion har studerats under lång tid. Idag behöver detektionsalgoritmen 6-7 pixlar i höjdlid för att upptäcka en människa. Vi behöver i fortsättningen studera vilken måligenkänningsprestanda är möjlig mot markmål. Med igenkänning (eng. Recognition) menar vi att målen automatiskt ska kunna delas in i mer detaljerade klasser såsom stridsvagn, lastbil, personbil, spaningsflygplan etc. För ett fordon ska man kunna bestämma formen och förhållandet mellan längd och bredd.

Inledande tester för att bedöma positioneringssystemets nuvarande precision och prestanda har gjorts och kommer fortgå under systemets vidareutveckling. I ett fortsatt arbete planeras också att kombinera/ersätta systemets stereokamera par med t.ex. en skannande laser (LIDAR). I fortsättningen bör även studeras hur hög positioneringsnoggrannhet som är möjlig mot markmål som fordon och människor.

6 Appendix

6.1 Detektering av rörliga mål från en plattform i rörelse

Här ges en mer teknisk beskrivning av de metoder som utvecklats i projektet för detektering av rörliga mål från en plattform i rörelse.

Videoföljning

Som beskrivits i avsnitt 2.2.2 har partikelpropagering (eng. *particle propagation*) använts för följning av punkter i bildplanet. Metoden baseras på beräkning av optiskt flöde mellan successiva bilder.

I projektet har två algoritmer för optiskt flöde studerats, dels Bouguets version av Lucas-Kanades klassiska algoritm [45], dels en mer modern ansats med totalvariationsbaserad regularisering och robust L^1 -norm för matchningsfelet [46]. Regulariseringen gör det i

princip möjligt att producera en 100 % tät skattning, dvs. en rörelsevektor för samtliga bildpunkter, även i omgivningar som saknar texturering; våra resultat visar dock att detta inte fungerar på ett stabilt sätt. Båda algoritmerna finns i funktionsbiblioteket OpenCV [47] och exekveras massivt parallellt på grafikkort, vilket möjliggör bearbetning i videotakt.

6.1.1 Luft till mark

Som beskrivits i avsnitt 2.2.2 föreslås för detektering av rörelse på marken från en flygande plattform end plan-parallaxansats baserad på bildstabilisering.

1. Stabilisera bakgrunden över en delsekvens av överlappande bilder:
 - a. Detektera ett plan i scenen
 - b. Ensa (registrera) successiva bilder mot planet
2. Detektera residualrörelse, dvs. punkter som efter stabiliseringen rör sig relativt planet.
3. Analysera och klassificera residualrörelsen:
 - a. Faktisk rörelse relativt den stationära bakgrunden
 - b. Parallaxrörelse för punkter i bakgrunden utanför det detekterade planet

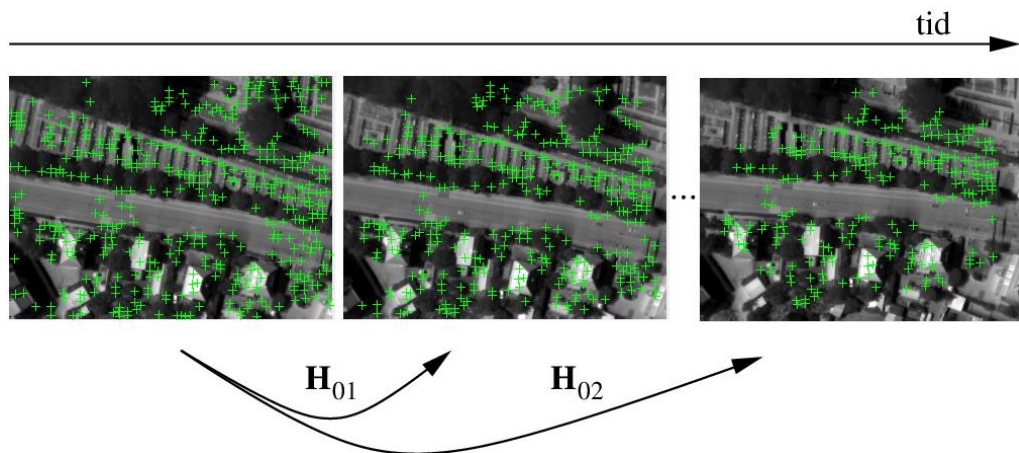
Bildstabiliseringen sker för successiva delsekvenser. I den första bilden i varje delsekvens identifieras ett antal intressepunkter, vilka sedan matchas mot motsvarande i nästa bild. Bland alla punktassociationer mellan de båda bilderna behålls den största delmängd som tillsammans rört sig på ett sätt som är konsistent med att de ligger i ett gemensamt plan i scenen. Villkoret är att punkterna avbildas med en s.k. homografi, vilket är en linjär transformation i homogena bildkoordinater. Om en punkt i bild m har koordinaterna

$\mathbf{p}_m^i = [x_m^i, y_m^i]^T$ är dess homogena koordinater $\tilde{\mathbf{p}}_m^i \equiv [x_m^i, y_m^i, 1]^T$. Homografien mellan referensbilden ($m=0$) och bild k ges då av uttrycket

$$c_k^i \tilde{\mathbf{p}}_k^i = \mathbf{H}_{0k} \tilde{\mathbf{p}}_0^i$$

Här är c_k^i en proportionalitetsfaktor och $\mathbf{H}_{0k}$ en 3×3 -matris. Vi använder tildesymbolen för att beteckna likhet upp till en skalär faktor, som i $\tilde{\mathbf{p}}_k^i \sim \mathbf{H}_{0k} \tilde{\mathbf{p}}_0^i$ för ovanstående uttryck.

RANSAC (*Random Sample Consensus* [50,54]) används för att finna den största delmängd punktassociationer som satisfierar homografivillkoret.



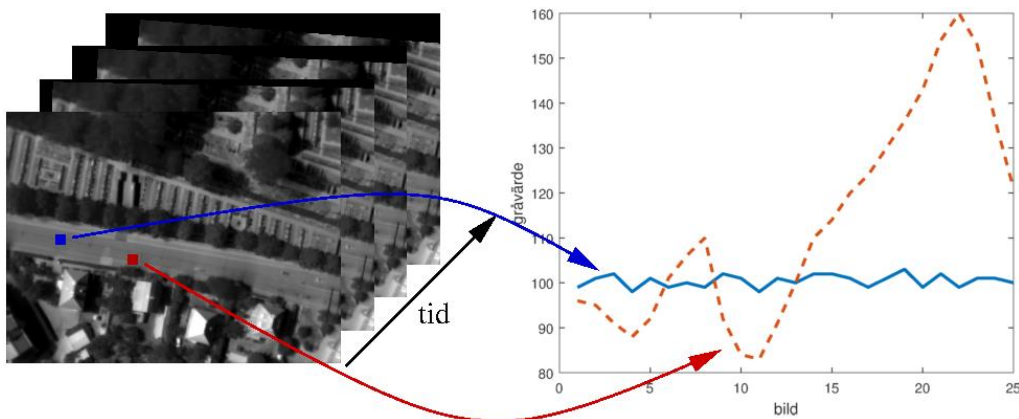
Figur 39. Identifiering av punkter i gemensamt plan. Koordinaterna i bildplanet för punkter vars rörelse är konsistenta med att de ligger i ett gemensamt plan i scenen transformeras enligt en homografi $\mathbf{p}_m^i = H_{0m}(\mathbf{p}_0^i)$ där $\mathbf{p}_0^i$ är positionen i referensbilden och $\mathbf{p}_m^i$ motsvarande i den m :te efterföljande bilden.



Figur 40. Stabilisering av delsekvens. Genom de skattade homografierna kan bilderna ensas (registreras) till den initiala referensbilden.

Punkterna följs sedan genom hela delsekvensen, och i varje steg används RANSAC för att identifiera de spår vars rörelse är konsistenta med ett plan, se figur 39. Eftersom plattformens rörelse typiskt är huvudsakligen parallell med markytan ställs allt högre krav på punkternas rörelse, vilket effektivt eliminerar punkter utanför det gemensamma planet. När samtliga bilder i delsekvensen bearbetats förfinas skattningen av homografierna $\mathbf{H}_{01}, \dots, \mathbf{H}_{0K}$ genom en iterativ optimeringsprocess kallad *bundle adjustment* [48,50].

Homografierna används sedan för att geometriskt registrera bilderna till den första referensbilden, se figur 40. Om bilderna nu läggs ovanpå varandra kommer gråvärdet för punkter i eller nära det skattade planet att variera lite i tiden (blå punkt/kurva i figur 41), medan gråvärdet för övriga punkter i allmänhet kommer att variera betydligt mera (exempelvis röd punkt/kurva, där ett objekt i rörelse passerar). I de punkter där variationen i tiden överstiger ett tröskelvärde startas ett spår, och partikelpropagering används för att följa den lokala omgivningen genom den stabiliserade sekvensen. Om två spår vid något tillfälle hamnar mycket nära varandra, förkastas det som rört sig kortast sträcka.



Figur 41. Skillnadsdetektion. Efter stabilisering analyseras variationen i gråvärde över tiden för varje bildpunkt. Punkter som avbildar strukturer i eller nära det skattade planet uppvisar typiskt en liten variation (blå punkt/kurva), medan värdet för övriga punkter varierar betydligt mera (t.ex. den röda punkten/kurvan, som är en bil i rörelse).



Figur 42. Klustring av närliggande spår i gemensam rörelse.

För att reducera mängden data för den fortsatta analysen grupperas spåren genom agglomerativ hierarkisk klustring [49]. Som avståndsmått mellan par av spår används det största euklidiska avståndet mellan samtidiga positioner. Närliggande spår med liknande rörelse genom sekvensen samlas till ett gemensamt objekt, vars centrumpunkt får representera samtliga spår i gruppen, se figur 42.

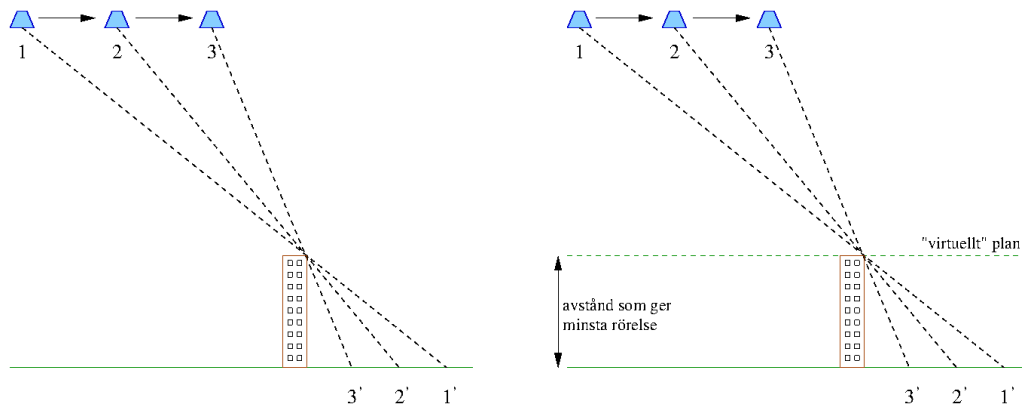
De objekt som är i faktisk rörelse relativt den stationära markbakgrunden utgör i allmänhet endast en delmängd av dem som identifierats genom ovan beskrivna procedur. Strukturer utanför referensplanet för stabilisering ger upphov till en parallaxrörelse. Situationen illustreras schematiskt i figur 43 (vänster). För att identifiera parallaxrörelser kan man resonera som i figur 43 (höger). Storleken på parallaxrörelsen beror av punktens vinkelräta avstånd från referensplanet. Det betyder att om stabiliseringen istället sker mot ett ”virtuellt” plan parallellt med referensplanet, men förskjutet till punktens position, uppstår ingen parallaxrörelse.

Om kamerans interna parametrar (fokallängd, principalpunkt, pixelstorlek, linsdistorsion, m.m.) är kända (eller kan bestämmas genom kalibrering) kan homografierna som används vid bildstabilisering uttryckas i normaliserade koordinater och skrivas [50]

$$\hat{\mathbf{H}}_{0k} = \mathbf{R}_{0k} \left(\mathbf{I} - \frac{\mathbf{t}_{0k}}{d} \hat{\mathbf{n}}^T \right), \quad k = 1, \dots, K$$

Här är $\hat{\mathbf{H}}_{0k} \equiv \mathbf{K} \mathbf{H}_{0k} \mathbf{K}^{-1}$, där $\mathbf{K}$ är kamerans kalibreringsmatris. $\mathbf{R}_{0k}$, en rotationsmatris, och $\mathbf{t}_{0k}$, en translationsvektor, beskriver kamerans rörelse mellan utgångspositionen och bild k . $\hat{\mathbf{n}}$ är referensplanets normal, och d det ortogonala avståndet mellan planet och kamerans utgångsposition. Det finns flera metoder för att bestämma de olika parametrarna givet en homografimatrix [51,52,53], men alla har gemensamt att det inte går att bestämma

storleken på translationsvektorn $\mathbf{t}_{0k}$, endast dess riktning och kvoten $\|\mathbf{t}_{0k}\|/d$. Det betyder alltså att det inte går att bestämma hur snabbt sensorn rör sig eller avståndet till referensplanet, endast förhållandet mellan dessa storheter, om inte ytterligare information från exempelvis en tröghetssensor finns tillgänglig. Dessutom finns för varje homografi alltid två möjliga lösningar för hela parameteruppsättningen. För att hitta den korrekta lösningen kan man beakta samtliga homografier i sekvensen. Dessa har en gemensam parameter, nämligen plannormalen $\hat{\mathbf{n}}$. Den korrekta lösningen fås då för varje homografi genom att kräva att man i samtliga fall får samma värde för normalvektorn; den alternativa lösningen är i allmänhet inte lika konsistent över tiden. Två viktiga specialfall av rörelse kan identifieras genom singularvärdesuppdelning av homografimatrisen. Om två singularvärden är lika rör sig kameran längs referensplanets normal, medan om samtliga tre värden är lika är translationen $\mathbf{t}_{0k} = \mathbf{0}$.



Figur 43. Identifiering av parallaxrörelse. Vänster: När kameran rör sig tycks byggnadens tak röra sig i den mot markplanet stabiliserade sekvensen. Höger: Om bilden istället stabiliseras mot ett plan parallellt med markplanet, men förskjutet till hustakets höjd, är taket fixerat genom sekvensen.

För att tillämpa den ovan föreslagna metodiken för identifiering av parallaxrörelse noterar vi att

$$\hat{\mathbf{H}}_{0k}(s) = \mathbf{R}_{0k}(\mathbf{I} - s \frac{\mathbf{t}_{0k} \hat{\mathbf{n}}^T}{d}), \quad k = 1, \dots, K$$

är en parametriserad familj av homografier för plan parallella med referensplanet. För varje punkt i parallaxrörelse finns då ett visst värde $s > 0$ sådant att om sekvensen stabiliseras genom registrering till motsvarande referensplan elimineras parallaxrörelsen fullständigt (bortsett från mätfel). Alternativt kan en jämförelse göras mellan den uppmätta parallaxrörelsen och en prediktion. I homogena koordinater ges prediktionen av

$$\tilde{\mathbf{p}}_{0k}^i(s) \sim \mathbf{K} \hat{\mathbf{H}}_{0k}(1)^{-1} \hat{\mathbf{H}}_{0k}(s) \mathbf{K}^{-1} \tilde{\mathbf{p}}_0^i = \tilde{\mathbf{p}}_0^i + (1-s) \frac{\hat{\mathbf{n}}^T \mathbf{K}^{-1} \tilde{\mathbf{p}}_0^i}{d - \hat{\mathbf{n}}^T \mathbf{t}_k} \mathbf{K} \mathbf{t}_k$$

Här är $\tilde{\mathbf{p}}_0^i$ en position i referensbilden och $\tilde{\mathbf{p}}_{0k}^i(s)$ den predikterade positionen i den geometriskt transformerade bild k . $\mathbf{K} \mathbf{t}_k$ är de homogena koordinaterna för projektionen av kameracentrum k in i referensbilden, den så kallade epipolen. Ekvationen säger därför att vid varje tidpunkt måste samtliga parallaxrörelser, uttryckta som skillnaden mellan utgångspunkten i referensbilden och den aktuella positionen i den geometriskt transformerade bilden, peka mot (eller från) en gemensam punkt i bilden. Punkter vars residualrörelse inte satisfierar detta villkor är i rörelse relativt den statistiska bakgrunden.

Vi kan verifiera att samtliga positioner längs ett spår i den stabiliserade sekvensen är konsistenta med parallaxrörelse genom att kontrollera att samma faktor s fungerar för att prediktera rörelsen. Exempelvis kan ett värde som minimerar de euklidiska avståndet mellan uppmätt och predikterad position bestämmas för den sista bilden i delsekvensen. Man kan sedan använda detta värde för s i uttrycket ovan för övriga bilder och kontrollera att predikteringen fungerar. Lämpligen beräknas s för flera bilder för att öka robustheten.

6.1.2 Mark till mark

I denna del av projektet har vi undersökt möjligheten att använda samtidig skattning av scenens 3D-struktur och sensorplattformens rörelse [50] för att detektera objekt i rörelse.

Metoden utgår ifrån följning av så många punkter som möjligt genom videosekvensen. De resulterande spåren används för att bestämma hur sensorplattformen rört sig (roterat och translaterat) relativt en utgångsposition. Denna information kan sedan utnyttjas för att skatta de följda punkternas position i scenen. Eftersom kamerans position vid varje ögonblick nu är känd kan de skattade punktpositionerna i rummet projiceras in i bildplanet och jämföras med de uppmätta spåren. Om skillnaden är stor kan spåret inte höra till den statiska bakgrunden.

Kamerans interna parametrar antas som tidigare vara kända. Detektionsprocessen börjar efter att de första spåren uppnått en viss minsta längd. Med spår som existerat mellan den första och den senaste bilden beräknas den så kallade *essentiella matrisen* $\mathbf{E}$ genom en RANSAC-baserad [54] robust skattningsmetod. $\mathbf{E}$ är en 3×3 -matris med egenskapen att

$$\tilde{\mathbf{p}}_2^T \mathbf{K}^{-T} \mathbf{E} \mathbf{K}^{-1} \tilde{\mathbf{p}}_1 = 0$$

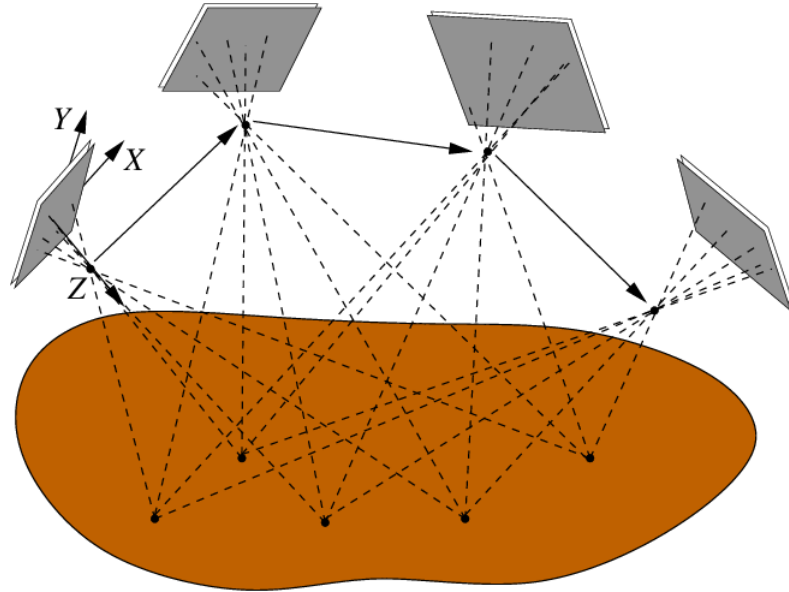
för samtliga punktkorrespondenser $\mathbf{p}_1 \leftrightarrow \mathbf{p}_2$ mellan två bilder som härrör från fixa objekt i scenen. Här är, som tidigare, $\mathbf{K}$ kamerakalibreringsmatrisen och $\tilde{\mathbf{p}} = (x, y, 1)^T$ homogena bildkoordinater. Man kan visa [50] att $\mathbf{E} \sim \mathbf{R}[\mathbf{t}]_x$, där $\mathbf{R}$ är rotationsmatrisen och $\mathbf{t}$ translationsvektorn för kamerans position i den andra bilden relativt den första. Här är

$$[\mathbf{t}]_x \equiv \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}$$

en matrisrepresentation av vektorprodukten, dvs. $[\mathbf{a}]_x \mathbf{b} = \mathbf{a} \times \mathbf{b}$. $\mathbf{R}$ och $\mathbf{t}$ kan extraheras från $\mathbf{E}$ genom singularvärdesuppdelning.

När kamerans förflyttning mellan två bilder nu är känd kan 3D-positionen för samtliga punktkorrespondenser mellan bilderna bestämmas genom *triangulering*. För samtliga bilder mellan den första och den senaste finns därefter en uppsättning korrespondenser mellan punkter i scenen med kända 3D-koordinater och deras projektioner i bildplanet. Att skatta en kameraposition givet dessa förutsättningar kallas att lösa *Perspective-n-Point (PnP)*-problemet, för vilket flera algoritmer är kända (exempelvis Fischler och Bolles ursprungliga tillämpning av RANSAC [54]). Skattningsarna av kamerapositioner och 3D-koordinater för följda punkter förfinas sedan genom *bundle adjustment* [48,50,55], i vilken man simultant skattar samtliga kamerapositioner och 3D-punktkoordinater genom att i minsta-kvadratmening iterativt minimera det genomsnittliga projektfelen i bildplanen, se figur 44. Detta ger ofta en mycket påtaglig förbättring av skattningsarnas kvalitet, men är beroende av att de initiala skattningsarna är i närheten av optimum. Det är värt att notera att avstånd endast kan bestämmas som multipler av det initialt uppmätta avståndet mellan den första och senaste kamerapositionen. För att bestämma absoluta avstånd krävs information från en annan källa, exempelvis en tröghetssensor.

I nästa bild uppdateras befintliga spår med en ny observation, och nya spår initieras där det är möjligt. En initial skattning av den nya kamerapositionen beräknas genom att lösa PnP-problemet med användning av de befintliga spår som fått en ny observation och som har en 3D-koordinat sedan tidigare. Därefter uppdateras 3D-koordinatskattningarna för samtliga aktiva spår genom triangulering mellan aktuell vy och den i vilket spåret initierats. Skattningarna av 3D-koordinater och kamerapositioner förfinas sedan åter genom *bundle adjustment*. För att uppnå beräkningsmässig effektivitet och kunna bearbeta videoströmmar i realtid delas denna upp i två delsteg: först uppdateras kamerapositionerna och 3D-koordinaterna samtidigt för en mindre delmängd spår, nämligen de som existerat längst tid, och därefter uppdateras 3D-koordinaterna för samtliga återstående spår givet de nya skattningarna för kamerapositionerna.



Figur 44. Bundle adjustment är en procedur där kamerans rörelse och 3D-positionerna för följda punkter skattas samtidigt genom att minimera det genomsnittliga avståndet mellan punkternas projektioner och de observerade punktkoordinaterna i bildplanen.

Spåren lagras i en cirkulärbuffert, vilket medför att gamla observationer så småningom skrivs över av nya. På samma sätt flyttas referenspositionen (koordinatsystemets origo) kontinuerligt ett steg framåt i tiden. Skattningen av 3D-koordinater och kamerapositioner sker alltså i ett tidsfönster relativt den aktuella tidpunkten.

Efter att spårens 3D-koordinater uppdaterats kan dessa projiceras in i de olika bildplanen i tidsbufferten och jämföras med de observerade positionerna. Till varje observation hör en skattning av mätningens osäkerhet (standardavvikelse) som antas proportionell mot förflyttningen sedan föregående bild. Om mätfelet antas normalfördelat och okorrelerat gäller att kvadratfelssumman över K bilder

$$\sum_{k=1}^K \frac{\|\mathbf{x}_k - \hat{\mathbf{x}}_k\|^2}{\sigma_k^2}$$

är chitvåfördelad med $2K$ frihetsgrader. Detta kan användas för att sätta en tröskel anpassad till en acceptabel falsklarmsnivå. Notera att, precis som i flygfallet, avsnitt 6.1.1, är det omöjligt att enbart genom geometrisk analys skilja parallaxrörelse från fallet när kamera och mål rör sig i samma eller motsatt riktning och med konstant relativ fart.

Algoritmen sammanfattas i figur 45.

Algoritm

1. Initiering:
 - a. Följ bildpunkter tills ett tillräckligt antal uppnått en minimilängd.
 - b. Skatta *essentiella matrisen* $\mathbf{E}$ mellan första och aktuell bild.
 - c. Extrahera från $\mathbf{E}$ rörelsen $(\mathbf{R}, \mathbf{t})$ mellan de båda bilderna.
 - d. Skatta de följda punkternas 3D-position genom *triangulering*.
 - e. Skatta rörelsen $(\mathbf{R}^k, \mathbf{t}^k)$ för samtliga bilder mellan den första och aktuella kamerapositionen, genom att lösa *PnP*-problemet.
 - f. Förfinas skattningarna av rörelse och 3D-punktpositioner genom *bundle adjustment*.
2. För varje ny bild:
 - a. Associera nya punktoobservationer till befintliga spår, eller starta nya.
 - b. Skatta aktuell kameraposition genom *PnP* för inmätta 3D-punkter.
 - c. Uppdatera 3D-positionen för samtliga spår genom *triangulering* mellan aktuell och första kameraposition.
 - d. Förfinas skattningar av rörelse och 3D-punktpositioner genom *bundle adjustment*.
 - e. Projicera 3D-punkter in i bildplanen och jämför med observerade spår. Stora avvikelser orsakas av punkter i rörelse eller associationsfel i spårbildningen.

Figur 45. Rörelsedetektion genom skattning av 3D-struktur och kamerarörelse.

7 Referenser

- [1] Degani, A. *Taming HAL: designing interfaces beyond 2001*, New York: Palgrave Macmillan, 2004.
- [2] Sarter, N. B., Woods, D. D. och Billings C. E. Automation surprises, G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics*, New York: Wiley, 1997.
- [3] Billings C. E., *Aviation Automation: The Search for a Human-Centered Approach*, Mahwah, NJ: Erlbaum, 1997.
- [4] Svenmarck, P. "Principles of human-robot coordination for improved manned-unmanned teaming", FOI Memo 1507, Linköping, 2005.
- [5] Karlholm, J. Implementation and evaluation of a method for detection of ground targets in aerial EO/IR imagery, FOI-R--1267--SE, Linköping, June 2004.
- [6] Bar-Shalom Y. och Li X. R., *Estimation and Tracking: Principles, Techniques, and Software*: Artech House, 1993.
- [7] Blackman S. S. och Popoli, R. *Design and Analysis of Modern Tracking Systems*: Artech House, 1999.
- [8] Kalman, R. E. "A New Approach to Linear Filtering and Prediction Problems", *Transactions of the ASME--Journal of Basic Engineering*, 82D, 1960.
- [9] Jazwinski, A. H. *Stochastic Processes and Filtering Theory*, Academic Press, 1970.
- [10] Julier, S. och Uhlmann, J. "Unscented filtering and nonlinear estimation", *Proceedings of the IEEE*, 92(3), mars 2004.
- [11] Gustafsson, F. och Hendeby, G. "Some relations between extended and unscented Kalman filters", *IEEE Transactions on Signal Processing*, 60(2):545-555, februari 2012. (Parallellpublicering: FOI-S--4014--SE.)
- [12] Gordon, N. J., Salmond, D. J. och Smith, A. F. M. "A novel approach to nonlinear/non-Gaussian Bayesian state estimation", *IEE Proceedings on Radar and Signal Processing*, 140, 107-113, 1993.
- [13] Doucet, A., de Freitas, N. och Gordon, N. *Sequential Monte Carlo Methods in Practice*: Springer Verlag, 2001.
- [14] Mahler, R. P. S. *Statistical Multisource-Multitarget Information Fusion*: Artech House, 2007.
- [15] Deleskog, V. och Hendeby, G. "Intelligent Surveillance with Multiple Sensor Platforms", *Proceedings of Reglermöte*, Linköping, juni 2014. (Parallellpublicering FOI-S--4729--SE)
- [16] Saha S., Hendeby G. och Gustafsson F. "Noise Adaptive Particle Filtering: A Hierarchical Perspective", *ISBA Regional Meeting and International Workshop/Conference on Bayesian Theory and Applications (IWCBA)*, 6-10 january 2013, Varanasi, Uttar Pradesh, India, 2013. (Parallellpublicering: FOI-S--4535--SE.)
- [17] Saha, S. och Hendeby, G. "Rao-Blackwellized particle filter for Markov modulated nonlinear dynamic systems", *Proceedings of IEEE Workshop on Statistical Signal Processing*, Gold Coast, Australien, juli 2014. (Parallellpublicering FOI-S--4758--SE.)
- [18] Saha, S., Hendeby, G. och Gustafsson, F. "Mixture Kalman Filters and Beyond" i *Current Trends in Bayesian Methodology with Applications*, 537-562, 2015. (Parallellpublicering: FOI-S--5199--SE.)
- [19] Roth, M., Hendeby, G. och Gustafsson, F. "EKF/UKF maneuvering target tracking using coordinated turn models with polar/Cartesian velocity", *Proceedings of 17th IEEE International Conference on Information Fusion*, Salamanca, Spanien, juli 2014. (Parallellpublicering: FOI-S--4753--SE.)
- [20] Veibäck, C., Hendeby, G. och Gustafsson, F. "Tracking of Dolphins in a Basin Using a Constrained Motion Model", *Proceedings of the 18th International Conference of Information Fusion*, Washington, DC, USA, 2015. (Parallellpublicering: FOI-S--5197--SE.)
- [21] Skoglund, M., Hendeby, G. och Axehill, D. "Extended Kalman Filter Modifications Based on an Optimization View Point", *18th International Conference of Information Fusion*, Washington, DC, USA, 2015. (Parallellpublicering: FOI-S--5197--SE.)

-
- [22] Roth, M., Fritsche, C., Hendeby, G. och Gustafsson, F. "The Ensemble Kalman Filter and its Relations to Other Nonlinear Filters", *Proceedings of the 2015 European Signal Processing Conference (EUSIPCO 2015)*, Nice, France, 2015. (Parallellpublicering: FOI-S--5198--SE.)
 - [23] Bleser, G., Damen, D., Behera, A., Hendeby, G., Mura, K., Miezal, M., Gee, A., Petersen, N., Mações, G., Domingues, H., Gorecky, D., Almeida, L., Mayol-Cuevas, W., Calway, A., Cohn, A. G., Hogg D. C. och Stricker, D. "Cognitive Learning, Monitoring and Assistance of Industrial Workflows Using Egocentric Sensor Networks", *PLoS ONE*, 10(6), 2015. (Parallellpublicering: FOI-S--5195--SE.)
 - [24] Reiss, A., Hendeby, G. och Stricker, D. "A novel confidence-based multiclass boosting algorithm for mobile physical activity monitoring", *Personal and Ubiquitous Computing*, 19(1): 105-121, 2015. (Parallellpublicering: FOI-S--5194--SE.)
 - [25] Bleser, G., Steffen, D., Reiss, A., Weber, M. och Hendeby, G., Fradet, L., "Personalized Physical Activity Monitoring Using Wearable Sensors", *Smart Health*, Lecture Notes in Computer Science, No. 8700, 99-124, 2015. (Parallellpublicering: FOI-S--5200--SE.)
 - [26] Durrant-Whyte, H. och Bailey, T. "Simultaneous localization and mapping: part I", *IEEE Robotics & Automation Magazine* 13(2), pp. 99-110, 2006.
 - [27] Bailey, T. och Durrant-Whyte, H. "Simultaneous localization and mapping (SLAM): Part II", *IEEE Robotics & Automation Magazine* 13(3), pp. 108-117, 2006.
 - [28] Rydell, J. och Emilsson, E. "Chameleon: Visual-inertial indoor navigation", *Proceedings of IEEE/ION PLANS*, 2012. (Parallellpublicering: FOI-S--4237--SE.)
 - [29] Rydell, J. och Emilsson, E. "Chameleon v2: Improved imaging-inertial indoor navigation," *Proceedings of the 26th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2013)*, 2013. (Parallellpublicering: FOI-S--4538--SE.)
 - [30] Emilsson, E. och Rydell J. "Chameleon on fire – thermal infrared indoor positioning", *Proceedings of IEEE/ION PLANS*, 2014. (Parallellpublicering: FOI-S--4715--SE.)
 - [31] Nyqvist, H. E., Skoglund, M. A., Hendeby, G. och Gustafsson, F., "Pose Estimation Using Monocular Vision and Inertial Sensors Aided with Ultra Wide Band", *2015 International Conference on Indoor Positioning and Indoor Navigation*, 2015. (Parallellpublicering: FOI-S--5207--SE.)
 - [32] Dellaert, F. "Factor Graphs and GTSAM: A Hands-on Introduction", Technical Report GT-RIM-CP&R-2012-002, Georgia Tech, 2012.
 - [33] Guivant, J., Nebot, E. och Baiker, S. "Autonomous navigation and map building using laser range sensors in outdoor applications", *Journal of Robotic Systems*, 17:3817–3822, 2000.
 - [34] Kaess, M., Ranganathan, A. och Dellaert, F. "iSAM: Incremental smoothing and mapping", *IEEE Transactions of Robotics*, 24(6):1365–1378, 2008.
 - [35] Teichner, W. H. "The Detection of Simple Visual Signal as a Function of Time on Watch". *Human Factors*, 16, 339-353, 1974.
 - [36] Helton, W. S., Dember, W. N., Warm, J. S. och Matthews, G. "Optimism, pessimism, and false failure feedback: Effects on vigilance performance". *Current psychology*, 18, 311-325, 2000.
 - [37] Irani, M. och Anandan, P. "A unified Approach to Moving Object detection in 2D and 3D Scenes", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(6), 577-589, 1998.
 - [38] Näsström, F., Bergström, D., Bissmarck, F., Grahn, P., Gustafsson, D. och Karlholm, J. "Prestandamått för sensorsystem", FOI-R--4139--SE, november 2015.
 - [39] Bernardin, K., och Stiefelwagen, R., "Evaluating multiple object tracking performance: the CLEAR MOT metrics", *EURASIP Journal on Image and Video Processing*, 2008.
 - [40] Schuhmacher, D., Vo, B.-T. och Vo, B.-N. "A consistent metric for performance evaluation of multi-object filters," *IEEE Transactions on Signal Processing*, 56(8), 3447–3457, 2008.

-
- [41] Ristic, B., Vo, B.-N., Clark, D. och Vo, B.-T. "A metric for performance evaluation of multi-target tracking algorithms", *IEEE Transactions on Signal Processing*, 59(7), 3452–3457, 2011.
- [42] Langhard, M. "Manöverkrigföring i svensk militärstrategisk doktrin", FHS, 2012, <http://www.diva-portal.org/smash/get/diva2:532075/FULLTEXT01.pdf>.
- [43] Beck, M. "Ledningsfilosofi i luftvärnet: Uppdragstaktik – Från decentralisering till centralisering", FHS, 2012, <http://fhs.diva-portal.org/smash/get/diva2:726045/FULLTEXT01.pdf>.
- [44] Näsström, F., Allvar, J., Berggren, J., Chevalier, T., Forsgren, R., Habberstad, H., Hedström, J., Hemström, F., Hendeby, G., Karlsson, M. och Persson, A., "Simuleringsbaserade metoder för sensorvärdering – Årsrapport 2011", FOI-R--3357--SE, december 2011.
- [45] Bouguet, J.-Y. "Pyramidal Implementation of the Lucas Kanade Feature Tracker. Description of the algorithm". Intel Corporation, Microprocessor Research Labs. OpenCV document. 2000.
- [46] Zach, C., Pock, T. och Bischof, H. "A Duality Based Approach for Realtime TV- L^1 Optical Flow. *Proceedings of the 29th DAGM conference on Pattern recognition*, 214-223, 2007.
- [47] OpenCV (open source computer vision). <http://opencv.org>. 2015-10-23.
- [48] Engels, C., Stewénius, H. och Nistér, D. "Bundle adjustment rules", *Photogrammetric Computer Vision*, 2006.
- [49] Hastie, T., Tibshirani, R., Friedman, J., *The Elements of Statistical Learning*, andra upplagan, Springer, 2009.
- [50] Hartley, R. och Zisserman, A. *Multiple view geometry in computer vision*, andra upplagan, Cambridge University Press, 2003.
- [51] Faugeras, O. och Lustman, F. "Motion and structure from motion in a piecewise planar environment", *International Journal of Pattern Recognition and Artificial Intelligence*, 2(3), 485-508, 1988.
- [52] Zhang, Z. och Hanson, A. R. "3D reconstruction based on homography mapping", *ARPA Image Understanding Workshop*, Palm Springs, CA, 1996.
- [53] Malis, E. och Vargas, M. "Deeper understanding of the homography decomposition for vision-based control". Forskningsrapport INRIA RR-6303, 2007.
- [54] Fischler, M. A. och 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.
- [55] Lourakis, M. I. A. och Argyros, A. A., "SBA: A Software Package for Generic Sparse Bundle Adjustment", *ACM Transactions on Mathematical Software*, 36(1), 1-30, 2009.

FOI är en huvudsakligen uppdragsfinansierad myndighet under Försvarsdepartementet. Kärnverksamheten är forskning, metod- och teknikutveckling till nytta för försvar och säkerhet. Organisationen har cirka 1000 anställda varav ungefär 800 är forskare. Detta gör organisationen till Sveriges största forskningsinstitut. FOI ger kunderna tillgång till ledande expertis inom ett stort antal tillämpningsområden såsom säkerhetspolitiska studier och analyser inom försvar och säkerhet, bedömning av olika typer av hot, system för ledning och hantering av kriser, skydd mot och hantering av farliga ämnen, IT-säkerhet och nya sensorers möjligheter.



FOI
Totalförsvarets forskningsinstitut
164 90 Stockholm

Tel: 08-55 50 30 00
Fax: 08-55 50 31 00

www.foi.se