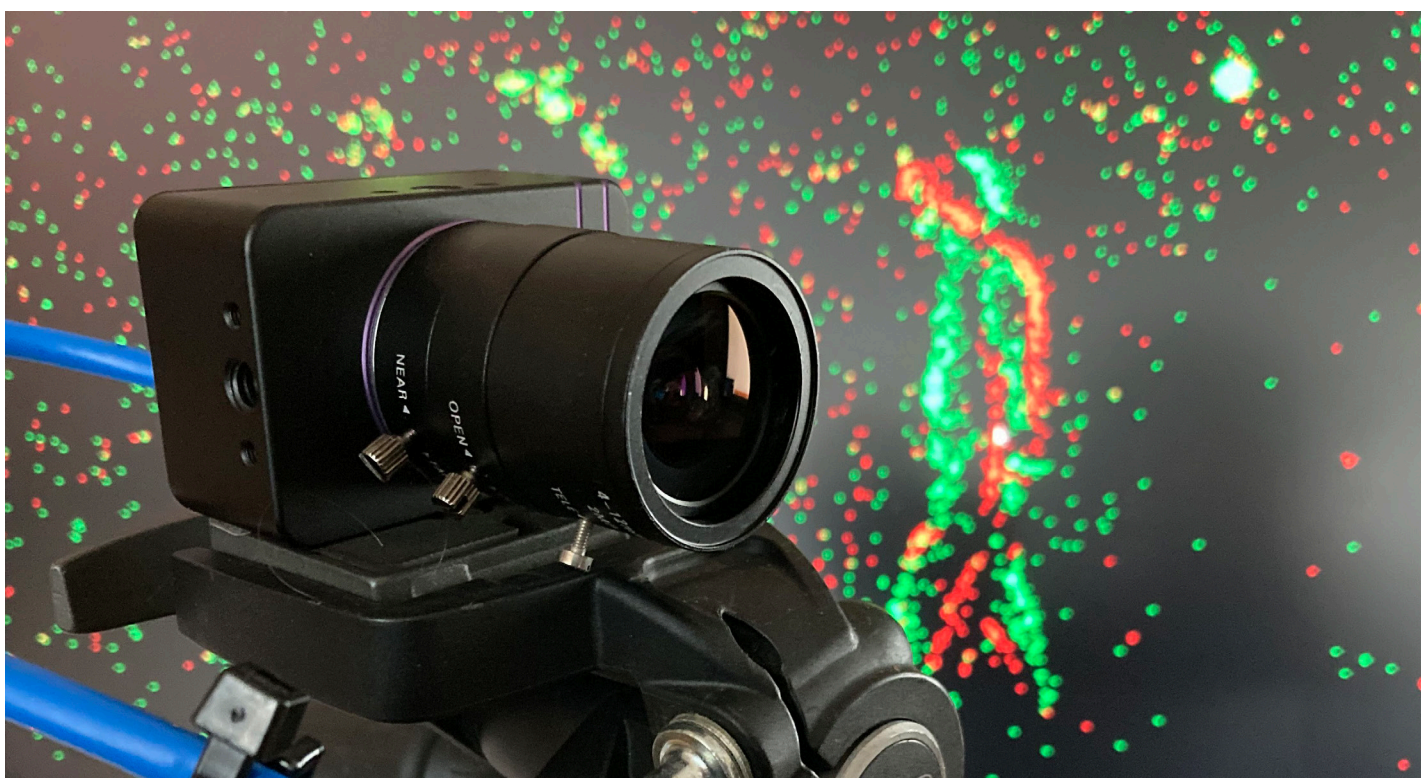


DAVID GUSTAFSSON, K-G STENBORG, HANNA HAMRELL,  
HANNES OVRÉN, JOHAN RIDEG, JENS GRUNDMARK



David Gustafsson, K-G Stenborg, Hanna Hamrell,  
Hannes Ovrén, Johan Rideg, Jens Grundmark

# Eventkamera för bevakning

Möjligheter och begränsningar

Titel	Eventkamera för bevakning – Möjligheter och begränsningar
Title	Event camera for surveillance - Capabilities and limitations
Rapportnr/Report no	FOI-R--5543--SE
Månad/Month	December
Utgivningsår/Year	2023
Antal sidor/Pages	35
ISSN	1650-1942
Uppdragsgivare/Client	FM Försvarsmakten
Forskningsområde	Sensorer och signaturanpassning
FoT-område	Sensorer och signaturanpassning
Projektnr/Project no	E716160
Godkänd av/Approved by	Emil Hjalmanson
Ansvarig avdelning	Cyberförsvar och ledningsteknik

Bild/Cover: K-G Stenborg, FOI

Detta verk är skyddat enligt lagen (1960:729) om upphovsrätt till litterära och konstnärliga verk, vilket bl.a. innebär att citering är tillåten i enlighet med vad som anges i 22 § i nämnd lag. För att använda verket på ett sätt som inte medges direkt av svensk lag krävs särskild överenskommelse.

This work is protected by the Swedish Act on Copyright in Literary and Artistic Works (1960:729). Citation is permitted in accordance with article 22 in said act. Any form of use that goes beyond what is permitted by Swedish copyright law, requires the written permission of FOI.

## Sammanfattning

Eventkameran är en ny typ av sensor där varje pixel oberoende av varandra genererar information om förändringar i ljusintensitet. Detta ger många fördelar jämfört med konventionella kameror så som en hög temporal upplösning, låg latens, högt dynamiskt omfång, låg bandbredd, samt låg strömförbrukning. De konventionella kameror som idag används för bevakning har en stor redundans i insamlad data då hela bilden uppdateras regelbundet, med relativt låg bildtakt, oberoende om förändring skett eller inte. Dessa har också svårt att fånga stora skillnader i ljusförhållande då alla bildens pixlar exponeras under samma tidsintervall.

Eventkamerans höga temporala upplösning gör att snabba förlopp som mynningsflammar kan urskiljas och frekvensanalys kan användas för att upptäcka propellrar och andra karaktäristiska frekvenser på objekt. Det höga dynamiska omfånget gör eventkameran robust mot överexponering och att det finns goda möjligheter att upptäcka saker av intresse trots lågljusförhållanden. Dess strömsnålhet är en fördel när övervakning bedrivs långt från existerande infrastruktur.

Idag finns eventkameror att köpa för överkomliga priser men de behöver kombineras med automatiska analysmetoder om dess data ska kunna användas för bevakningstillämpningar. Forskningen kring specifika analysmetoder är mycket aktiv och nya lovande resultat presenteras kontinuerligt. I denna rapport presenteras några möjliga tillämpningar för eventkameran inom bevakningsområdet.

Nyckelord: bevakning, övervakning, eventkamera, neuromorfisk sensor

## Summary

The event camera is a new type of sensor where each pixel independently generates information about changes in light intensity. This provides many advantages compared to conventional cameras such as a high temporal resolution, low latency, high dynamic range, low bandwidth, and low power consumption. Conventional cameras currently used in surveillance applications have a large redundancy in collected data as the whole image is updated regularly, at a relatively low rate, regardless of whether there is new information. They also struggle with large variations in lighting conditions as all the pixels in the image share the same exposure time.

The high temporal resolution of the event camera enables the capture of fast processes such as muzzle flares, and frequency analysis methods can be used to detect propellers and other characteristic frequencies on objects. The high dynamic range makes the event camera robust to overexposure and allows operation in low light conditions. The low power consumption is an advantage when conducting surveillance far from existing infrastructure.

Today, event cameras have affordable prices, but they need to be combined with automatic analysis methods to be used for surveillance applications. In this report, we describes some possible surveillance applications for the event camera.

Keywords: surveillance, event camera, neuromorphic sensor

## Innehåll

<b>1</b>	<b>Inledning .....</b>	<b>7</b>
<b>2</b>	<b>Eventkameran.....</b>	<b>9</b>
	2.1 Grundprinciper .....	9
	2.2 Sensorhårdvara .....	11
	2.3 Sensorns utveckling och framtid.....	11
	2.4 Kameraspecifikationer .....	12
	2.5 Styrkor och svagheter.....	13
<b>3</b>	<b>Eventbaserad signalbehandling.....</b>	<b>16</b>
	3.1 Bildbaserad bearbetning.....	16
	3.2 Eventbaserad bearbetning .....	16
	3.3 Neuromorfisk beräkningshårdvara .....	17
<b>4</b>	<b>Eventkamera för bevakningstillämpningar .....</b>	<b>18</b>
	4.1 Övervakning genom bakgrundsmodellering.....	18
	4.2 Utmanande ljusförhållanden .....	20
	4.3 Upptäckt och observation av små mål .....	21
	4.4 Frekvensanalys av en scen .....	22
	4.5 Vibrometri.....	22
	4.6 Snabba förlopp .....	23
	4.7 Bildrekonstruktion .....	25
	4.8 Rörliga övervakningskameror.....	27
	4.9 Djupinformation genom fokussvep .....	28
	4.10 Strömsnål övervakning .....	29
<b>5</b>	<b>Analys och slutsatser .....</b>	<b>31</b>
	5.1 TRL-nivå .....	31
	5.2 Värdering .....	32
<b>6</b>	<b>Referenser.....</b>	<b>33</b>
<b>7</b>	<b>Bilaga: Händelser som önskas upptäckas vid bevakning .....</b>	<b>35</b>



# 1 Inledning

Kameror används ofta inom bevakning för att i realtid stötta en operatör med bildmaterial, eller för lagring av bildmaterial för senare användning i utredningar eller som bevismaterial. Som stöd för operatören används i vissa fall även olika typer av algoritmer för detektion, klassificering och målföljning av exempelvis fordon och personer.

Den typ av konventionella kameror som ofta används vid övervakning avbildar en scen som en serie ögonblicksbilder insamlade vid diskreta tidpunkter definierade av en förbestämd och regelbunden (synkron) bildtakt. Varje bild förmedlar information (ljusintensitet) från samtliga pixlar, oavsett om informationen har ändrats något sedan föregående bild eller inte. Under antagandet att bilden i sig inte är slutmålet med insamlingen, utan att endast betrakta som mätdata för fortsatt analys och tolkning, är en sådan insamlingsteknik ineffektiv. Samtidigt skapas en stor mängd onödig redundans hos inspelat data. Med konventionell kamerateknik krävs efteranalys av (hela) bilderna för att kunna upptäcka rörelser och andra förändringar i scenen vilket kräver både energi och beräkningskapacitet. Den begränsade temporala upplösningen och att det förlöper en viss tid mellan varje ny bildtagning i bildsekvensen gör dessutom att viktig information trots den stora datamängden riskerar att missas.

Eventkameran, en så kallad *neuromorfisk sensor*<sup>1</sup>, är en relativt ny sensorteknik där varje pixel istället individuellt och asynkront genererar signaler, så kallade *event* eller *spikar*, när den uppmätta ljusintensitetsförändringen överstiger ett tröskelvärde. En eventsignal innehåller information om vilken pixel som genererat det, tiden då det skedde (eventet skapades), samt huruvida ljusintensiteten ökat eller minskat. Den här tekniken har många fördelar jämfört med konventionella kameror bland annat tack vare att signalen är gles, sensorn har hög temporal upplösning och att den är extremt strömsnål. Med gles menas i detta fall att signal endast uppstår *när* (temporalt) och *där* (spatialt) det finns aktivitet i scenen. Tekniken har därför stor potential att användas inom ramen för snabba och billiga system för övervakning. I Figur 1 och Figur 2 visas kort skillnaden mellan en eventkameran data och bilder från en konventionell kamera, mer om detta i nästa kapitel.

Kapitel 2 förklarar teknikens grundprinciper, dess styrkor och svagheter samt vilka aktörer som är med och utvecklar sensorn. För att kunna utnyttja eventkamerors unika egenskaper krävs anpassad signalbehandling och i kapitel 3 beskrivs läget för utveckling av sådan teknik. Eventkameran skulle kunna ha tillämpningar inom bevakning och i kapitel 4 beskrivs olika sätt på vilka sensorn skulle kunna användas för detta. Rapporten avslutas i kapitel 5 med sammanfattande analys och slutsatser kring teknikens möjligheter inom området.

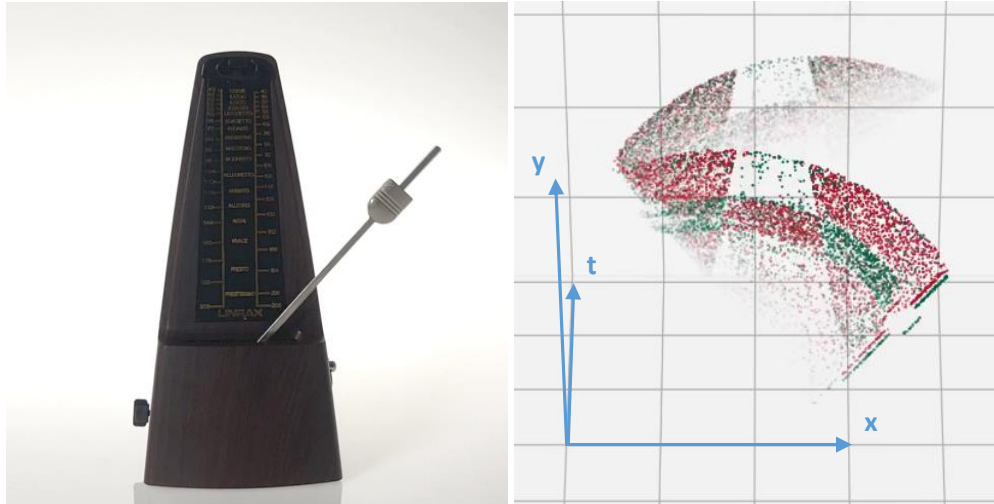
Denna rapport har tagits fram inom forskningsprojekten *Avancerade spaningsensorer*<sup>2</sup> och *Händelsebaserad avbildning – Försvarstillämpningar* samt en direktbeställning från Försvarmakten.

---

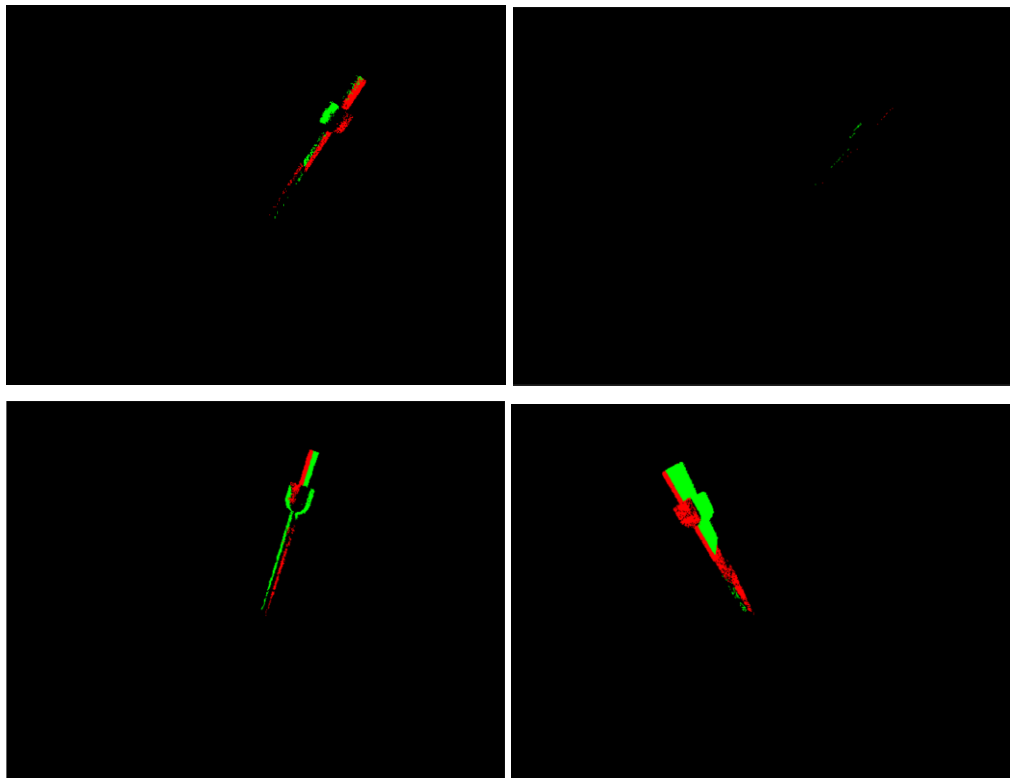
<sup>1</sup> En neuromorfisk visuell sensor har som mål att efterlikna hur biologiska ögon fungerar.

<sup>2</sup> Ett projekt inom Försvarmaktens samlingsbeställning för forskning och teknikutveckling inom Sensorer och signaturanpassning (FoT SoS, AT.9220423).





Figur 1 En metronom med en pendel som rör sig fram och tillbaka. Vid aktuell tidpunkt byter pendeln precis rörelseriktning från att svänga mot höger till att börja svänga mot vänster. Till vänster en bild från en konventionell videokamera. Till höger en 3D-illustration av uppmätt eventdata med tiden  $t$  som den tredje axeln. Äldre events försvinner bort i  $t$ -led och blir dessutom med tiden transparenta i denna beskrivning av det spatiotemporal rummet. Gröna punkter motsvarar en intensitetsökning där pixeln blivit ljusare och röda punkter motsvarar en intensitetssänkning i pixlarna. Ljussättningen gör att pendeln vänstra kant är ljusare än bakgrunden, medan övriga pendeln är mörkare än bakgrunden. När pendeln rör sig mot höger blir pixlarna mörkare när pendeln passerat (dvs röda i denna illustration) och när pendeln rör sig mot höger blir pixlarna efter pendeln ljusare (dvs gröna i denna illustration). Bilder från videon i presentationen av [1].



Figur 2 Fyra visualiseringar från eventdata på metronomen med pendeln i rörelse. Eventdata illustreras ofta förenklat som bilder med data insamlat under ett tidsintervall, i detta fall omkring 1/25 sekund. I svarta pixlar på bilden har inga events skett under detta tidsintervall, i gröna pixlar var det senaste eventen under tidsintervallet en intensitetsökning och röda pixlar betyder att senaste eventen var en intensitetssänkning. Uppe till vänster rör sig pendeln mot höger med röda pixlar när pendelns mörka högra sida rör sig över ljus bakgrund. Uppe till höger har pendeln nått höger ändläge och står under tidsintervallet i det närmaste still vilket ger få antal events. I bilden nere till vänster har pendeln börjat röra sig mot vänster. Nere till höger rör sig pendeln fortfarande mot vänster men har nått högre hastighet så den har rört sig längre under det givna tidsintervallet, jämfört med föregående bild.

## 2 Eventkameran

En eventkamera skiljer sig mycket från konventionella kameror. För att förstå hur eventkameran skulle kunna användas inom bevakning behöver man därför först förstå tekniken. Avsnitt 2.1 förklarar grundprinciperna för sensorn och avsnitt 2.2 förklarar hur hårdvaran fungerar. I avsnitt 2.3 presenteras kort de aktörer som varit med och utvecklat sensorn och exempel på den utveckling som pågår. Några av de nuvarande sensorerna på marknaden och deras specifikationer presenteras i avsnitt 2.4 och avslutningsvis listas några av för- och nackdelarna med sensorn i avsnitt 2.5.

### 2.1 Grundprinciper

Eventkameror är tidsregelbundna (asynkrona) sensorer, där varje pixel oberoende av varandra genererar en signal eller ett event ("händelse")<sup>3</sup> vid en intensitetsförändring hos det infallande ljuset. Varje event beskrivs av en tidpunkt, en position och en polaritet (ökning eller minskning) hos intensitetsförändringen. Utsignalen från en eventkamera blir därmed en asynkron sekvens av events där volymen av genererad data bestäms av aktuell dynamik i scenen – ju fler förändringar i scenen desto fler events. Denna princip är inspirerad av hur biologiska synsystem fungerar och ett synonymt begrepp som ofta används är neuromorfisk avbildning (eng. *neuromorphic imaging*). Eventkameran ger alltså inte hela intensitetsinformationen av det infallande ljuset på pixlarna som en konventionell kamera, utan genererar bara information om intensitetsförändringar.

Ett exempel på hur data från ett event kan se ut visas i Figur 3, där en eventkamera av typen DAVIS346 fått en intensitetsökning i en pixel vid en specifik tidpunkt.

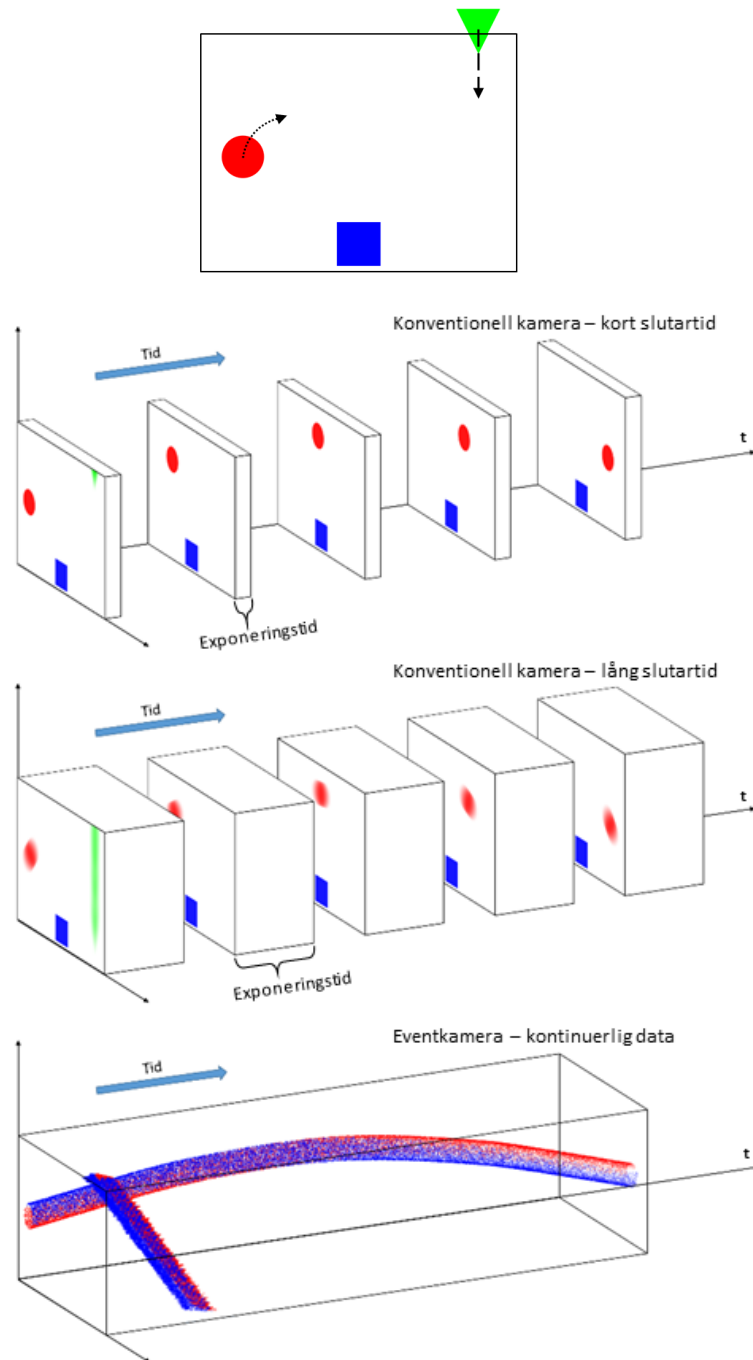
```
x: 59
y: 24
timestamp: 1540823063185872941
polarity: True
```

Figur 3 Ett exempel på ett event från en DAVIS346-kamera. Polariteten "True" betyder att detta event är för en intensitetsökning som skett i pixeln (59,24) vid tidpunkten 2018-10-29 14:24:23.185872941. De tre sista decimalerna är utanför kamerans temporala upplösning (som är ca. en mikrosekund).

Den asynkrona dataströmmen från en eventkamera kan visualiseras på olika sätt. För att kunna visas i tryckt form är det vanligt att man aggregerar eventdata från en kort tidsperiod för att presenteras som en bild. En standardmetod som används i den här rapporten är att räkna antalet positiva och negativa event för varje pixel och placera dessa i röd- respektive grönkanalen i en vanlig färgbild. I Figur 1 och Figur 2 visar på två olika sätt att visualisera eventdata, dels som en aggregerad bild enligt ovan men även som punkter i ett spatiotemporalt 3D-rum. Trots att data ofta presenteras i form av en bild är det viktigt att påminna sig om att en eventkamera alltså inte genererar bilder utan en asynkron ström av intensitetsförändringar per pixel.

En illustrerad jämförelse av data från en konventionell kamera och en eventkamera finns i Figur 4. Två bildsekvenser från konventionella kameror med olika slutartid jämförs med en eventkameran data, för samma händelseförlopp i en scen. Den konventionella kameran bygger upp diskreta bilder insamlade under specifika exponeringstider medan eventkameran samlar data i det närmaste kontinuerligt under hela tidsförloppet.

<sup>3</sup> I denna rapport använder vi ordet "events" för de små pixelförändringarna och "händelser" för skeenden och situationer i scenen. I andra texter kan dock "händelser" användas även för de små intensitetsförändringarna på pixelnivå.

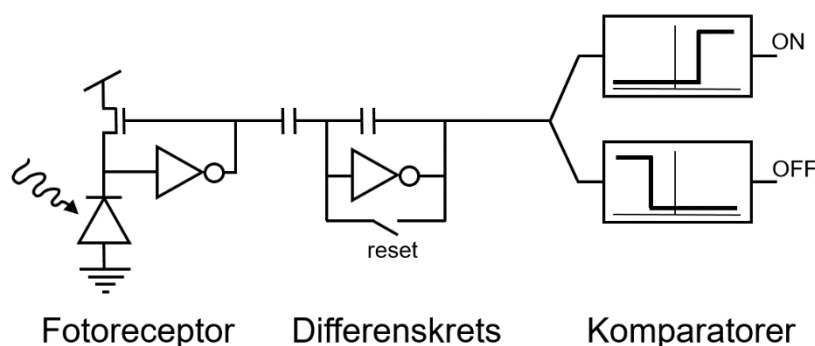


Figur 4 En scen med tre objekt som har olika händelseförlopp, samt en statisk vit bakgrund. Den blå kvadraten är helt statisk, den röda cirkeln har en relativt långsam kaströrelse och den gröna triangeln har en väldigt snabb fallande rörelse, med start delvis utanför synfältet, i början av tidsintervallet. Högst upp illustreras scenen som den beskrivs från kamerornas synfält. Rad två illustrerar en konventionell kamera som fångar objektens position vid specifika diskreta tidsintervall. I detta fall har kameran en kort slutartid vilket gör att det är stora delar av tiden då ingen exponering av bilder sker. Den gröna triangeln hinner börja komma in i scenen under den första bilden i sekvensen men har redan hunnit passera innan nästa bild börjar samlas in. Den långsamma röda cirkeln får en rätt skarp bildtakt, men en längre slutartid som gör att större andel av tiden används för exponering. Detta leder till att en större del av rörelsen från den gröna triangeln samlas in i den första bilden, fast med kraftig rörelseoskärpa. På liknande sätt blir den röda cirkeln mer utsuddad av rörelsen under exponeringstiden. I den sista bilden visas insamlad (i det närmaste) kontinuerlig eventdata, i denna illustration genererat med ESIM [2]. För detta fall blir intensitetssänkningar (när en pixel blir mörkare) blåa punkter och intensitetsökningar röda punkter i det spatiotemporal rummet. Den gröna triangelns hela rörelse samlas in i början av förloppet och den röda cirkelns hela kastbana syns tydlig. Däremot ger den statiska blå kvadraten (och bakgrunden) inga events. Färgerna på de rörliga objekten kan ej eller urskiljas eftersom eventsensorn är monokrom.

I det visuella våglängdsområdet är registrerad ljusintensitet i en kamerapixel en produkt av belysningen av scenen (illuminansen) och reflektansen hos de material och objekt i scenen som pixeln avbildar. Över kortare tid är belysningen i naturliga scener att betrakta som relativt konstant. Uppmätta intensitetsförändringar signalerar därför att sensorn har rört på sig i förhållande till scenen, att ett objekt i scenen har rört på sig i förhållande till sensorn eller att objektet i sig självt har förändrats. Till exempel, för en stillastående kamera som betraktar en helt statisk scen skulle det därför inte genereras några events alls, fränsett eventuellt brus. Då finns ingen ny information om scenen att registrera och kameran behöver därför inte skicka data.

## 2.2 Sensorhårdvara

Exakt hårdvaruimplementering kan skilja sig något åt mellan tillverkarna av eventkameror. I en implementation (från Inivation) används s.k. DVS-pixelchip, där DVS står för *Dynamic Vision Sensor*. En DVS-pixel består av tre huvudkomponenter (se Figur 5); en fotodiod med en spänningsstabiliserande förstärkare som skapar en signal som är proportionell mot logaritmen av infallande ljusintensitet, en differenskrets med förstärkning av signalvariationer och borttagning av absolutnivå (DC-komponent) samt två komparatorer som detekterar positiva respektive negativa signalförändringar och genererar digitala events när dessa överskrider en specificerad tröskelnivå.



Figur 5 Förenklat kretsschema för en DVS-pixel, med fotoreceptor, differenskrets med återställare samt komparatorer.

På grund av logaritmisk respons har tekniken en hög dynamik, typiskt 120 dB, vilket kan jämföras med 55-60 dB hos konventionell CMOS-teknik. Konventionella kameror har därför vanligen problem att hantera scener med stort dynamiskt omfång där vissa regioner i bilden blir mörka (underexponerade) medan andra regioner i bilden blir mättade (överexponerade).

## 2.3 Sensorns utveckling och framtid

Genom EU-projektet CAVIAR (2002–2006) kunde den första praktiskt användbara eventkameran demonstreras. Detta skedde år 2006 av Lichtsteiner med flera vid ETH (Eidgenössische Technische Hochschule) i Zürich, Schweiz, och tekniken benämndes då som Dynamic Vision Sensor (DVS). Grundprinciperna för just DVS beskrivs i avsnittet ovan. Idag säljs DVS-kameror främst via schweiziska företaget Inivation med huvudkontor i Zürich. Senare års utveckling har främst fokuserat på reduktion av pixelstorlek och ökning av pixelupplösningen. Under 2019 inledde företaget ett samarbete med Samsung.

Huvudkonkurrent till Inivation är det franska företaget Prophesee, med huvudkontor i Paris. Prophesee inledde 2020 ett samarbete med Sony.

Den amerikanska försvarsforskningsmyndigheten DARPA presenterade 2021 ett fyraårigt samarbetsprojekt mellan Raytheon, BAE Systems och Northrop Grumman som går under namnet FENCE (Fast Event-based Neuromorphic Camera and Electronics). Projektet ska ta fram IR-baserade eventkameror med asynkrona utläsningskretsar och separata signalbehandlingslager för analys av spatiala och temporala signaler. Ambitionen är att hålla total sensoreffekt under 1,5 W. Möjliga militära tillämpningar som nämns för tekniken är autonoma fordon, robotik och IRST (InfraRed Search and Track).

Inom NATO planeras att 2024 genomföra en studie, *exploratory team* [3], för att planera ett informationsutbyte, *research task group*, som ska studera tillämpningar av eventkameror: *Applications for Neuromorphic Cameras*.

Sammantaget är tekniken förhållandevis ung, men ett flertal aktörer är med och driver på utvecklingen [4]. Framförallt verkar det som att utvecklingen går mot kameror med högre pixelupplösning och ökad tillverkning av eventkameror som är IR-baserade.

## 2.4 Kameraspecifikationer

Det finns i nuläget (våren 2023) ett flertal olika eventkameror kommersiellt tillgängliga och deras specifikationer återfinns i Tabell 1. Till detta kan tilläggas att företaget Inivation erbjuder flera varianter av eventkameror, som alla bygger på likartad sensorhårdvara:

- DVXplorer lite, DVXplorer samt DVXplorer mini: Dessa varianter kan enbart generera event.
- DAVIS346: generar event och bild från samma sensor-matris
- DVXplorer Duo: en eventsensor kombinerat med en högupplöslig visuell kamera.

Utöver enskilda eventkameror erbjuder Inivation även en stereo-rigg med två stycken eventkameror som kan användas för 3D-rekonstruktion av snabba förlopp.

Tabell 1: Specifikationer för sensorer som finns på marknaden våren 2023.

Namn	DAVIS346	DVXplorer lite	DVXplorer	DVXplorer mini	DVXplorer – Duo	SilkyEvCam (VGA)	SilkyEvCam (HD)
Tillverkare	Inivation	Inivation	Inivation	Inivation	Inivation	Prophesee	Prophesee
Spatial upplösning	346 x 260	320 x 240	640 x 480	640 x 480	640 x 480	640 x 480	1280 x 720
Temporal upplösning	1 us	65- 200 us				200 us	100us
Max genomströmning (throughput)	12 MEPS	100 MEPS	165 MEPS	450 MEPS	450 MEPS	50 MEPS	20 MEPS
Latens (svarstid)	<1 ms	<1 ms	<1 ms	<1 ms	<1 ms		
Dynamiskt omfång	Approx. 120 dB	Approx. 90 dB				>120 dB	>120 dB
Kontrastkänslighet	14.3% (on), 22.5% (off) (with 50% pixel response)	13% (with 50% of pixels respond), 27.5% (with 99.9% of pixels respond)					
Pixel pitch	18.5 um	18 um	9 um	9 um	9 um		
Bild	346 x 260	Ingen bild genereras från dessa sensorer			Full HD	Ingen bild	Ingen bild
Storlek (W x H x D) mm	40 x 60 x 25	40 x 60 x 25	40 x 60 x 25	29 x 29 x 32	32 x 80 x 92	30 x 30 x 36	30 x 30 x 36
Vikt	100 g	75 g	100 g	43 g	220 g	40 g	42 g
Strömförbrukning	< 180 mA	<140 mA	<140 mA	<140 mA	<140 mA	300 mA	100 mA

FOI har undersökt tre olika eventkameror – en DVXplorer och en DAVIS346 från Inivation samt en SilkyEvCam (VGA-upplösning) som bygger på teknik från Prophesee, se Figur 6. FOI har också ett utvecklingskit för Speck från Synsense vilken beskrivs i avsnitt 3.3.



Figur 6 Fyra eventkameror och en storleksreferens. Till vänster DVXplorer och DAVIS346 (som både samlar in events och vanlig bild på samma gång) från Inivation. I mitten SilkyEvCam från Prophesee. Slutligen ett Speck utvecklingskit från Synsense.

## 2.5 Styrkor och svagheter

Eventkameror reagerar asynkront på ljusförändringar och får därmed fördelar i förhållande till konventionell teknik. En fördel är snabbhet med låg fördröjning (latens) mellan event och utsignal. Enskilda event kan tidsstämplas med mikrosekundupplösning och latensen är under 1 millisekund. Detta skapar möjligheter att upptäcka och följa mycket snabba händelseförlopp och reducerar risk för rörelseoskärpa. Det bör dock noteras att en eventkamera, liksom alla sensorer, har en begränsad bandbredd för dataöverföring. Jämfört med konventionella kameror finns dock potential att reducera bandbreddskrav hos transmissionslänkar, särskilt om scenen har låg eller ingen temporal dynamik.

Den högsta förändringstakt som kan detekteras ökar med ljusintensiteten och således kommer en ljusstark scen kunna avbildas med högre tidsupplösning än en ljussvag eller mörk.

En ytterligare styrka med tekniken är den låga effektförbrukningen på detektornivå, som typiskt är 10  $\mu$ W till 10 mW beroende på antalet genererade events. Inbäddade kamerasytem, med inbyggd signalbehandling, har uppvisat förbrukning under 100 mW vilket ska jämföras med konventionella bildsensorer av CMOS-typ som vanligtvis ligger på omkring 1 W. Den låga förbrukningen möjliggör sensorintegration på mindre plattformar och/eller i system som kräver hög uthållighet.

Den höga dynamiken från sensorns logaritmiska respons skapar högre robusthet mot utfrätning eller bländning från starka ljuskällor. Den möjliggör också avbildning av scener som innehåller mycket ljusstarka och ljussvaga partier samtidigt.

Nackdelar med eventbaserad avbildning är till stor del kopplad till den låga mognadsgraden hos tekniken. Eftersom utsignalen från en eventkamera skiljer sig väsentligt från den signal som erhålls från konventionell kamerateknik så krävs specialiserad signalbehandling, som delvis är utvecklad idag. Ett exempel är teknikens ökade känslighet för egenrörelse, brus och andra dynamiska effekter, vilket kräver specialanpassade filter. Avsaknaden av standardiserade testmetoder gör det dessutom svårt att jämföra prestanda mellan olika eventkameror, men European Machine Vision Association (EMVA) har meddelat att man arbetar på att ta fram rekommendationer. För tillfället är eventkameror dyrare på grund av högre tillverkningskostnader, vilket sannolikt kommer kunna pressas vid en ökad massproduktion av tekniken. Pixelupplösningar och fyllnadsfaktorer är i dagsläget också väsentligt sämre, men integrering av tekniken i modern bakupplyst och stackad CMOS-teknik har förbättrat prestandan under senare år.

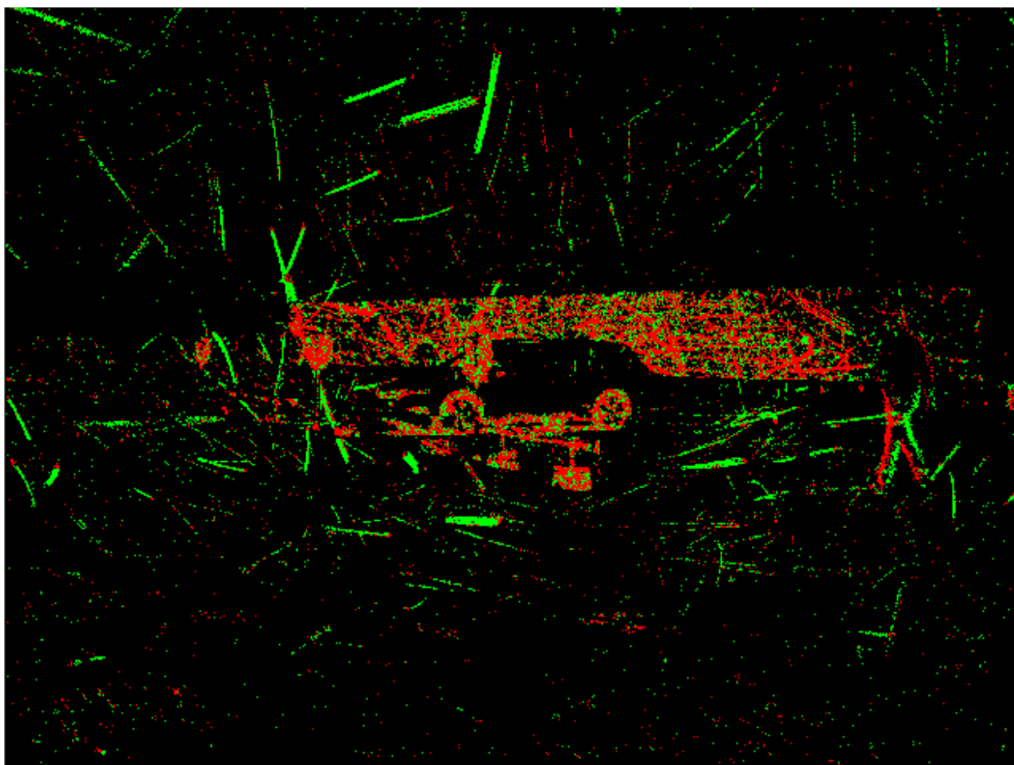
Vad gäller sensorns förmåga att hantera olika typer av väderfall så kan snö och regn ge upphov till extra mycket event då bakgrunden inte längre är statisk. Detta kan, på samma sätt som för konventionella kameror, minska möjligheterna att upptäcka små objekt på stora avstånd. Större objekt vanliga vid ett övervakningsscenario bör ändå kunna detekteras i dessa typer av scenarier så länge som det finns tillräckligt stor kontrast mellan objektet av intresse och bakgrunden. Ett exempel på sensorns förmåga att hantera snöfall

syns i Figur 7 där de ljusa snöflingorna ger upphov till extra stor ljusintensitetsförändring när de passerar framför mörkare bakgrunder. Därför syns extra många event mot de delar av bakgrunden som var mörka, till exempel däckets på den vita bilen i mitten av scenen.

Eventkamerans styrkor kan sammanfattas med följande punkter. De har bland annat studerats på FOI av Hamrell et al. i [5].

- **Hög temporal upplösning** – eventkamerans temporala upplösning är mycket hög och även aktiviteter med begränsad temporal utbredning kan upptäckas och analyseras. Vid övervakning är det få händelser som är så snabba att en eventsensor inte skulle generera events.
- **Låg bandbredd** – eventkameror genererar bara data när något händer i scenen och då endast för de pixlar där något har inträffat. I relativt statiska scener som ofta är fallet vid långtidsövervakning kommer sensorn att generera relativt lite data och i ideal fallet av en helt statisk scen genereras ingen data alls.
- **Låg strömförbrukning** – eventkameran genererar bara data när något inträffar i scenen. För statiska scener med få eller inga förändringar genereras således inga data och sensorns strömförbrukning blir därmed låg.
- **Stort dynamiskt omfång** – då sensorn använder sig av en logaritmisk skala, saknar global slutare och varje pixel genererar event oberoende av varandra blir det dynamiska omfånget för sensorn mycket stort.
- **Spatial upplösning** – jämfört med vissa sensorer inom ett par likvärdiga tillämpningsområden, som PIR-sensorer, har eventkameran en spatial upplösning som gör den mer användbar. Dagens konventionella HD-kameror har dock en högre upplösning än de flesta kommersiellt tillgängliga eventkameror. Fast eventkamerornas spatials förmåga är ofta tillräcklig för att tydligt urskilja former på rörliga objekt, vilket gör att de kan användas för målklassificering och kanske även måligenkänning. När tekniken mognar och blir mer kommersiell kommer eventkamerornas upplösning dessutom troligtvis att öka.

Eventkamerans svagheter kan sammanfattas med följande punkter.



Figur 7 Bild under intensivt snöfall skapad av insamlade events under 1/50 sekund. Gröna pixlar betyder att en intensitetsökning skett senast i pixeln och röda pixlar betyder att senaste eventen där var en intensitetsökning. Snöflingor nära eventkameran ger tydliga svansar där de yr fram och är synliga oberoende av bakgrund. En mörk husvägg finns i mitten av bildens höjded och där genereras många events eftersom det är hög kontrast mellan vita snöflingor och mörk vägg. Däremot är det lite kontrast mellan snöflingorna och den ljusa himlen samt snötäckta marken, där ger den fallande snön för dessa delar av scenen upphov till färre events. I mitten av bilden står en vit bil parkerad med få events på karossen. Dess svarta däck syns tydligt eftersom kontrasten jämfört med de fallande snöflingorna där är störst. En person med ljus tröja och mörka byxor syns gående i bildens högra sida.



## 3 Eventbaserad signalbehandling

För att dra nytta av eventkamerans unika egenskaper krävs tillämpningsspecifik signalbehandling och hårdvara som effektivt och energisnålt kan bearbeta asynkrona dataströmmar med varierande, och ibland mycket hög, dataakt. Metoder för behandling av eventdata kan delas in i två huvudkategorier, bildbaserad bearbetning och eventbaserad bearbetning.

### 3.1 Bildbaserad bearbetning

Ett sätt att hantera eventdata är att göra om det till en sekvens av tvådimensionella bilder baserat på en spatiotemporal omgivning. Denna typ av representation har fördelen att flertalet redan existerande bildbehandlingsalgoritmer kan användas mer eller mindre rakt av, även om bilderna skiljer sig en del i karaktäristik från bilder fångade med konventionella kameror.

Ett vanligt exempel är att låta bilden representera ett tvådimensionellt histogram, där antalet event som inträffat under ett visst tidsintervall sparas för varje pixel. I en annan variant sparas istället tidstämpeln för det senaste eventet i varje pixel. Denna representation kallas "tidsyta" (eng. *time surface*) och inkodar rörelsehistoriken i bilden så att pixlar med högre intensitet indikerar att de uppdaterats senare än pixlar med lägre intensitet.

Många bildbehandlingsalgoritmer fungerar relativt bra på båda typerna av eventbaserade bildrepresentationer som diskuteras ovan. Kantinformation, som vanligen extraheras från event och explicit återfinns i flertalet representationer, är ofta central vid bildanalys. Ett hörn i en scen betraktad med en eventkamera under rörelse, där en uppsättning av event representeras som en tidsyta, får då en specifik signatur som kan detekteras och följas på ett effektivt sätt. Kantinformation är också central vid beräkning av optiskt flöde (eng. *optical flow*), vilket är en viktig grundteknik för objektdetektion och som kan beräknas utifrån beskrivna representationer.

Nackdelen med att representera eventdata som bilder är att man inte utnyttjar glesheten i eventströmmen och att signalbehandlingen därmed kan kräva onödigt mycket beräkningsresurser. Dessutom tappar man informationen i varje events tidstämpling i relation till övriga närliggande event.

### 3.2 Eventbaserad bearbetning

Ett alternativ till bildbaserad bearbetning är eventbaserad bearbetning. Antingen behandlas då enskilda event när de kommer eller en mindre uppsättning event från en liten spatiotemporal omgivning. Varje event, eller uppsättning av event, analyseras tillsammans med information som sparats från tidigare analyserade event för att extrahera önskad information.

Genom att tolka tidsdimensionen som en tredje rumsdimension så kan eventdata representeras som ett 3D-punktmoln och man kan använda sig av metoder anpassade för detta. Exempelvis kommer en linje som rör sig över bildplanet ge upphov till ett plan i det spatiotemporalala 3D-rummet vilket gör att denna kan detekteras med hjälp av 3D-metoder.

Andra metoder för att hantera grupper av event är att liksom i föregående avsnitt skapa bildliknande representationer, men med ett större antal kanaler som bättre bevarar all information som finns i eventdata. För en människa är de här "bilderna" inte nödvändigtvis tolkningsbara alls men däremot har den här typen av representation visat sig fungera bra tillsammans med standardmetoder för maskininlärning.

För asynkron bearbetning av eventdata, det vill säga att man bearbetar enskilda event direkt när de uppstår, så är de två dominerande teknikerna filtrering och *spikande neurala*

*nätverk*. Ett exempel på filtrering är särdragsextraktorn Arc\* [6] som uppdaterar en tidsyta för varje inkommande event, för att därefter avgöra om det nyss inkomna eventet representerar ett särdrag eller inte.

Artificiella neurala nätverk (ANN) av den typ som oftast används inom bildbehandling är en förenklad matematisk modell av hur den mänskliga hjärnan bearbetar information. Spikande neurala nätverk [7] kan sägas vara ytterligare ett steg närmare den mänskliga hjärnan där den stora skillnaden ligger i hur information representeras och bearbetas. Där ett ANN arbetar på bilder så består datarepresentationen för ett SNN av spikar i tid vilket är princip samma sorts data vi får från en eventkamera. Neuroner i ett SNN tränas till att bli känsliga för vissa mönster av spikningar, både temporalt och spatalt. När ett sådant mönster uppstår genererar neuronerna själv en spik som den skickar till de i nätverket nedströms inkopplade neuronerna. Den stora fördelen med SNN är att databearbetningen sker *golest* både temporalt och spatalt – beräkningar sker bara när något inträffar och där (i bildplanet) det inträffar. Detta ger större möjligheter att dra nytta av eventsensorns fördelar, som strömsnålhet, hög dynamik och höga tidsupplösning.

### 3.3 Neuromorfisk beräkningshårdvara

Fördelen med den låga strömförbrukningen hos eventkameran går förlorad om databearbetningen inte också är energieffektiv. Konventionell beräkningshårdvara för ANN ligger typiskt mellan 10-300W vilket är flera magnituder större än de milliwatt som är typiska för en eventkamera. Genom att kombinera eventkameran med en beräkningsenhet som är bättre anpassad för eventdata kan mycket vinnas i energiförbrukning. Det finns idag flera exempel på *neuromorfisk beräkningshårdvara* som implementerar ett SNN i hårdvara, och kommersiella produkter i olika utvecklingsstadier finns från exempelvis Intel (USA), Brainchip (Australien) och Synsense (Schweiz/Kina).

Ett nämnvärt exempel är *Speck* från Synsense som kombinerar en eventkamera (128x128 pixlar) med en neuromorfisk beräkningsenhet. I ett demoexempel har Speck kunnat klassificera gester (vinka, luftgitarr, etc.) med en rapporterad total strömförbrukning runt 20mW för både sensorn och klassificeringen.

## 4 Eventkamera för bevakningstillämpningar

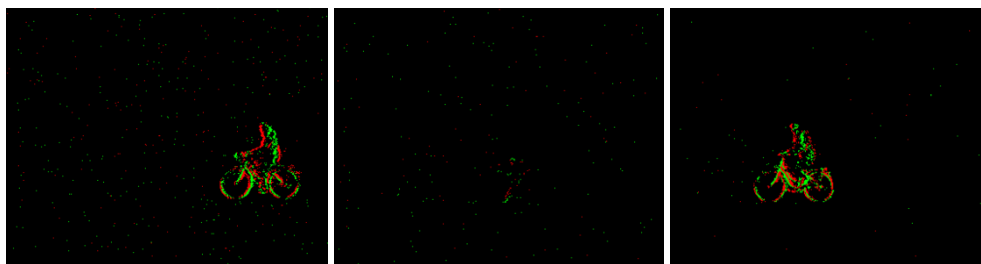
En av Försvarsmaktens uppgifter är att bevaka sina anläggningar och för detta kan stöd från sensorsystem behövas. I detta kapitel utvärderas tekniken mot några av de händelser som man typiskt vill upptäcka, händelserna beskrivs i Bilaga 1. De flesta tillämpningar och tekniker som beskrivs i kapitlet har på någon nivå undersökts av FOI, medan ett par är mer spekulativa men finns med då de passar in i den övergripande beskrivningen och lämpar sig för bevakningssituationer.

### 4.1 Övervakning genom bakgrundsmodellering

I ett bevakningsfall där en eventkamera är placerad för att övervaka ett område, såsom en väg eller en begränsad yta nära ett skyddsvärt objekt, så påminner den bildinformation som direkt genereras av eventkameran till stor del om den bildinformation som kan skapas med en konventionell övervakningskamera tillsammans med signalbehandlingsverktyg för så kallad bakgrundsmodellering.

Med bakgrund menas den ”kuliss” av föremål och element som är statiska för scenen, till exempel byggnader, vegetation, lyktstolpar etcetera. Med förgrund menas de övriga objekt och element som inte tillhör bakgrunden men som för tillfället ändå kan ses i scenen, till exempel personer eller fordon under färd. Bakgrundsmodellering innebär att etablera en matematisk beskrivning över vilka bildelement som tillhör scenens bakgrund. Mot en sådan bakgrundsmodell kan man sedan upptäcka förändringar och andra anomalier i ett kontinuerligt flöde av bilder. Man kan upptäcka när personer eller fordon kommer in i scenen eller när fordon som tidigare fanns i en scen försvinner ut ur scenen.

Bakgrundsmodellering är en vanligt förekommande metod för att analysera och extrahera information ur videosignaler från övervakningssituationer. Denna modellering är således ett viktigt (för-)steg i många bildbehandlingsprocedurer. Samtidigt är det ofta svårt och beräkningskrävande att etablera en robust modellering med konventionella kameror. Eventkameran utgör ett enklare och mer resurseffektivt alternativ till detta.



Figur 8 Detaljer från tre bilder som skapats av aggregerade events. En cyklist blir synlig i synfältet (vänster), stannar, hoppar av cykeln och lämnar en väska (mitten) och hoppar på cykeln och cyklar vidare (höger). Under rörelsen genereras event som möjliggör detektion och klassificering av objekt och aktiviteter. Då rörelsen upphör blir objekten snabbt en del av bakgrunden och därmed svåra att upptäcka och klassificera.

Informationen som extraheras vid bakgrundsmodellering påminner om informationen som fås genom att aggregera events från en eventkamera, men med mycket lägre temporal upplösning. Detta eftersom en konventionell kamera har en relativt låg bildtakt jämfört med den kontinuerliga strömmen av events från eventkameran, vilket gör att information från bakgrundsmodelleringen får samma låga uppdateringstakt som den konventionella kameran varifrån dess data härstammar.

Flera av problemen med informationsextraktion med bakgrundsmodellering är också direkt överförbara till information extraherade från eventströmmar såsom:

- Objekt smälter med tiden in i vad som uppfattas som bakgrund – personer och objekt i scenen som inte genererar event blir en del av bakgrunden och dessa blir därmed svårare att upptäcka. Objekt, så som väskor eller fordon, som lämnas i en scen kommer efter en stund (alternativt momentant) att ingå i modellen av bakgrunden och blir därmed inte observerbara i eventströmmen. Figur 8 visar en sekvens av bilder med en cyklist. I den första och sista bilden är cykeln i rörelse och genererar därför event. I den mittersta bilden är cykeln stilla och den blir därför en del av bakgrunden. Detta kan möjligen lösas med hjälp av bildrekonstruktion, se avsnitt 4.7.
- Falsklarm – objekt som kontinuerligt förändras i scenen genom rörelse eller andra orsaker kommer generera event och potentiellt falsklarm. Rörliga objekt såsom trädgrenar och flaggor kommer generera event, men även mindre uppenbara snabba förändringar så som lysrör som blinkar med en viss frekvens.

Ovan nämnda problemen kan åtgärdas genom metoder som styr tidskonstanterna för när stillastående saker ska börja ingå i bakgrunden samt filtrering av bakgrundsklotter så som blinkande ljus eller vajande grenar. Detta leder dock till att ett mer resurskrävande steg läggs till i signalbehandlingskedjan för data från eventkameran, och det måste tas ställning till om det tillför nödvändig information.



Figur 9 Detalj i bilder av en cyklist på väg att stanna upp i scenen i två olika ljusförhållanden. Visuella bilder till vänster och motsvarande tidpunkts eventdata till höger. Den statiska bakgrunden i scenen genererar inga events utan det är den rörliga cyklisten som framträder i eventdata. I den mörka scenen är antalet events färre än i dagsljus, men i övrigt innehåller de liknande data.

## 4.2 Utmanande ljusförhållanden

Det är ofta av stor vikt att kunna detektera och klassificera objekt, även under utmanande ljusförhållanden. Eventkameror har en logaritmisk intensitetsskala och med detta ett stort dynamiskt omfång. Sensorn har därför både en viss mörkerförmåga och kan användas för att analysera händelser i starka ljuskällor såsom strålkastare, flammor, ljusbågar och dylikt. Den stora dynamiken möjliggör även att sensorn kan användas i scener där det samtidigt förekommer mycket svagt belysta bildelement och starka ljuskällor.

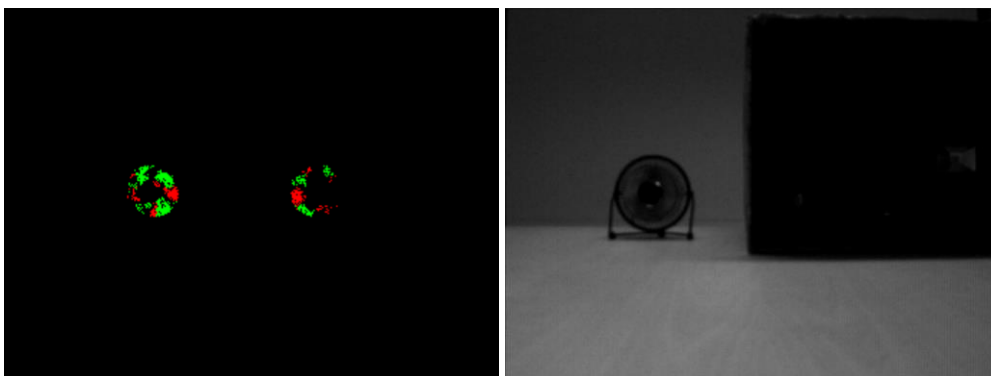
Eventkameran genererar likartade events för samma aktivitet under olika ljusförhållanden. Om en person går eller ett fordon åker igenom en scen en mörk kväll, en mulen dag eller en solig dag kommer likartade events att genereras. De events som en eventkamera genererar är alltså relativt oberoende av ljusförhållandena i scenen.



Figur 10 En bil med påslagna strålkastare passerar genom scenen i två olika ljusförhållanden. Visuella bilder till vänster och motsvarande tidpunkts eventdata till höger. Den statiska bakgrunden i scenen genererar inga events utan det är mest bilens konturer som framträder i eventdatat. Dock gör strålkastarna i samverkan med bilens rörelse att omgivningen lysas upp varierande, vilket genererar fler events i mörkerfallet.

I Figur 9 och Figur 10 visas hur en scen uppfattas av en videokamera jämfört med en eventkamera, i det här fallet en DVXplorer, vid olika grader av mörker. Notera hur de rörliga lamporna (bildens strålkastare) påverkar eventdatat i mörkerfallet för Figur 10, jämfört med den mer likartade eventdatan för cyklisten i Figur 9.

I Figur 11 visas eventdata från två likadana snurrande fläktar med skillnaden att den ena befinner sig upplyst och den andra i kraftig skugga. En jämförelse med DAVIS346-kamerans visuella bild indikerar eventkamerans förmåga att uppfatta rörelser i mörker.



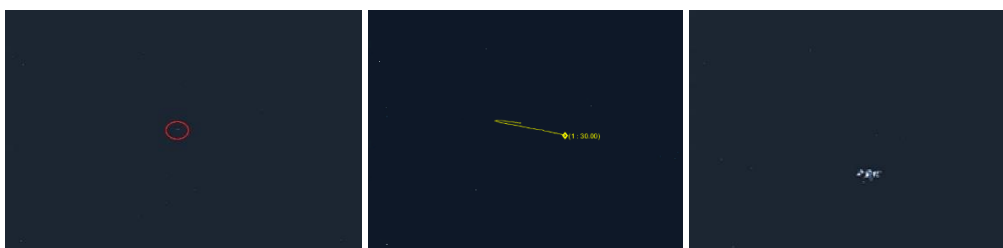
Figur 11: Två snurrande fläktar med den högra inne i en ej upplyst låda. Scenen filmad med en DAVIS346-kamera och till vänster visas dess events och till höger dess visuella bild. I eventen syns bladens rotation tydligt även om det är färre event vid den högra fläkten. I den visuella bilden syns inte den högra fläkten trots att en lång slutartid används så att videon bara har 0,8 bilder per sekund.

### 4.3 Upptäckt och observation av små mål

Upptäckt av små mål som endast täcker en eller ett fåtal pixlar är ofta en utmaning med konventionella kameror. Det finns goda möjligheter att upptäcka små rörliga mål i relativt statiska scener med en eventkamera under såväl lågljusförhållanden som mer ljusintensiva förhållanden. Den statiska bakgrunden genererar få eller inga event emedan målobjektet genererar en kontinuerlig ström av event. Dessa förändringar framträder tydligt vid visuell inspektion av eventdata, eller kan användas i kombination med signalbehandling för automatisk måldetektion och -följning.

Vid detektering och följning av rörliga mål bestående av ett fåtal pixlar är brusnivån i sensordata, speciellt under lågljusförhållanden, ett problem. Brus gör att eventkameran genererar event relativt slumpmässigt. Bruskänsligheter varierar för olika pixlar och vissa överkänsliga pixlar genererar regelbundet falska event orsakade av brus som kan vara svåra att separera från event genererade från verkliga mål. I alla fall utan att man filtrerar bort överkänsliga pixlar [8]. Om målets rörelse kan modelleras, med hjälp av klassiska metoder för målföljning, kan uppsättningar av event aggregeras ihop till målspar och falska detektioner kan filtreras bort.

Den högre temporal upplösningen gör också att målföljning bör förenklas då det är kort tid mellan detektioner i sekvensen som ska associeras. Vid eventbaserad bearbetning av eventdata är detektionerna i det närmaste kontinuerliga jämfört med de tydligt diskreta detektionerna från en konventionell kameras bildsekvens. Om det finns flera mål i varandras närhet bör dessutom den höga temporal upplösningen göra att det är betydligt enklare att särskilja mellan de olika målen under associationssteget.



Figur 12 UAV-detektering och följning på olika avstånd. I detta eventdata är alla events representerade av vita pixlar oberoende av om de kommer från ökad eller minskad intensitet. Till vänster visas en UAV som täcker två pixlar på 150 meter och i mitten visas målspar för den två pixlar stora UAV:n under en minut, där detektering och följning gjorts i eventdatat. Till höger visas samma UAV på kortare avstånd, där detaljer framträder som skulle kunna användas för typbestämning.

Detektion och följning av UAV:er på större avstånd är ett område av stort intresse där eventkamera har stor potential. I Figur 12 visas event för en UAV på ca 150 meters avstånd och som i bildplanet täcker ca 3 pixlar. UAV:n framträder relativt tydligt i sensordata trots den begränsade utbredningen och den höga brusnivån. En uppsättning event kan aggregeras ihop till ett målspar för UAV:n vilket visas i mittersta bilden i Figur 12. Högerbilden i Figur 12 visar samma UAV på 30 meters avstånd då fler detaljer blir synliga vilket öppnar upp för typbestämning av farkosten. Ett möjligt stöd vid detektion och klassificering kan fås med hjälp av frekvensanalys av data, se avsnitt 4.4.

## 4.4 Frekvensanalys av en scen

Många scener innehåller aktiviteter som är periodiska på olika sätt. Dessa periodiciteter kan potentiellt användas för filtrering av eventströmmen. Sådan filtrering är möjlig, men samtidigt är det ännu oprövat hur väl det fungerar för bevakningstillämpningar.

Realtidsfrekvensanalys på eventdata är ett mycket starkt verktyg som öppnar upp för många tillämpningar. Detta studeras bland annat i FOI:s artikel av Grundmark et al. i [1] där blinkade LED-lampor och rotationshastigheter från propellrar behandlas.

Det finns flera olika scenarier där frekvensanalys kan ha stor nytta på eventdata:

- Bortfiltrering av ointressant information – belysning som blinkar med en viss frekvens, grenar som vajar med en bestämd rörelse eller vågor med en viss frekvens ger upphov till onödiga event. Sådana repetitiva rörelser kan filtreras bort i vanliga kameror för att underlätta detektion av små rörelser från exempelvis personer som kort tittar fram, rör ett sikte eller ett vapen, samt personer som simmar bland vågorna. Eventkamerans höga temporala upplösning bör potentiellt underlätta sådan filtrering och möjlighet att särskilja det som är intressant från det repetitiva.
- Klassificering utifrån frekvensinnehåll – vid detektion och målföljning kan frekvensinnehåll associeras till målobjektet och användas för särskiljning och typbestämning. Olika UAV-modeller ger upphov till olika frekvenser genom sina propellrar som potentiellt kan användas för särskiljning. Vissa sensorer har svårt att särskilja fåglar från UAV:er, men på frekvensnivå bör det vara stor skillnad mellan vingslag och propellrar.

Frekvensanalys av eventströmmen bör med andra ord både kunna användas för att särskilja intressant och ointressant information, men även för att direkt upptäcka och klassificera objekt genom deras i förväg kända karaktäristiska frekvenser.

## 4.5 Vibrometri

Ljudvågor orsakade av till exempel tal får närliggande lättrorliga objekt i scenen att vibrera. På liknande sätt kan objekt som fordon med påslagen motor skapa ljud och vibrationer i karossen. Dessa vibrationer kan uppfattas av snabba och känsliga sensorer så som lasersystem och höghastighetskameror.

På liknande sätt kan potentiellt eventkameradata användas för att studera frekvensinnehållet i pixlarna och detta användas för avlyssning eller för att upptäcka specifika högfrekventa ljud. Ett exempel på det sistnämnda skulle kunna handla om att upptäcka om fordon har motorn igång och även avgöra fordonstyp. Detta skulle kunna vara av intresse i ett bevakningsscenario där man även passivt vill kunna upptäcka olika typer av ljud eller klassificera till exempel fordon, utan att använda sig av fler sensorer och med en relativt billig och enkel installation. Sådan information kan både vara av intresse att utvinna direkt under pågående händelse, men även under en senare forensisk undersökning av scenen. Användning av eventkamera för denna tillämpning är i nuläget inte utvärderad.

## 4.6 Snabba förlopp

Som visades med den gröna triangeln i Figur 4 är eventkameran väldigt bra på att fånga hastiga rörelser, där en konventionell kamera riskerar att missa dem i tidsspannet mellan bilderna i sekvensen eller att rörelseoskärpa försvårar möjligheten att urskilja förloppet. Förutom rörelser av objekt finns det andra snabba förlopp som explosioner eller mynningsflammar som är väldigt viktiga att registrera. Vi har tidigare talat om höga frekvenser och mer upprepande rörelser, medan detta avsnitt behandlar snabba förlopp som sker en eller ett begränsat antal gånger. Förlopp som önskas upptäckas eller undersökas mer detaljerat, både temporalt och spatialt.

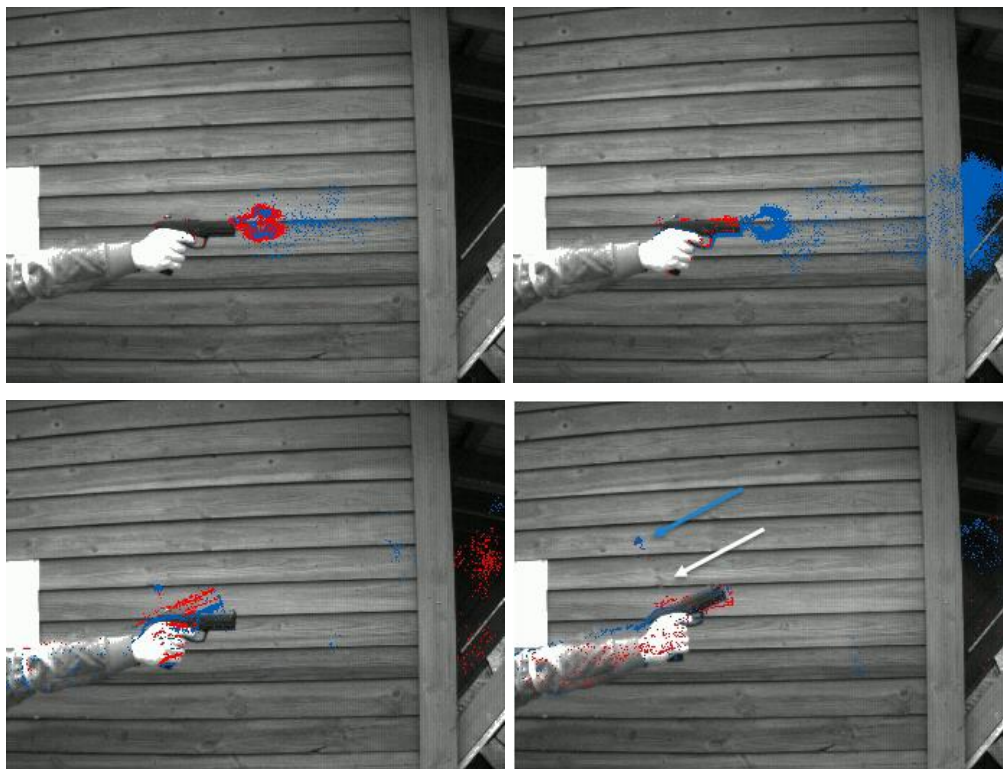
I Figur 13 visas två efterföljande visuella monokroma videobilder med en bildtakt på 25 bilder/sekund och en exponeringstid på 7,5 ms från DAVIS346-kameran. Detta betyder att det är 32,5 ms mellan varje exponering av bild i videoströmmen, det vill säga kameran samlar bara in visuell bild under omkring 19% av tiden. Dessa bilder är tagna vid avfyringen av pistolen men det går ej att urskilja på bilderna, förutom rekylen som fått handen att röra sig. En undersökning av tidkoderna för visuella bilder och eventdata tyder på att events från mynningsflamman börjar drygt 1,5 ms efter att exponeringen av den första av de visuella bilderna upphört och att nästkommande visuella bild börjar exponeras åtminstone 9 ms efter att mynningsflamman upphört att synas i eventströmmen.



Figur 13 Avfyrning av pistol på skjutbana. Visuella bilder från DAVIS346-kameran. Detta är de två bilderna som är närmast skottet i videoströmmen. Den första bilden lite innan avfyringen och den andra efter när rekylen redan har flyttat handen.

I Figur 14 visas aggregerad eventdata från samma kamera med en av oss vald bildtakt på 1500 bilder/sekund, och den är överlagrad på närmast föregående tidsstämplade visuella bild från sensorn. Eftersom det är olika bildtakt på gråskalebilderna och vår visualisering av eventdata, betyder det på att det går 60 eventdatabilder innan den visuella bilden i bakgrunden ändras. När en bakgrundsbild ändras är de första 12 bilderna med eventdata tagna under exponeringen av den visuella bakgrundsbilden och de efterföljande 48 bilderna innehåller events från tidsspannen då ingen visuell exponering sker.





Figur 14 Events överlagrade på senaste visuella bild. Båda typerna av data från DAVIS346-kameran. Här är eventen visualiserade med en bildtakt på 1500 bilder/sekund. Överst till vänster är bild nummer 17 med den bakgrundsbilden, då mynningsflamman verkar vara som mest intensiv. Överst till höger är bild nummer 24 med samma bakgrundsbild, då rök och eventuellt tryckvåg syns en bit framför pistolen. Nere till vänster är bild nummer 58 med fortfarande samma bakgrundsbild och där syns tydligt rekyl i eventen. Slutligen nere till höger är bild 60 från den efterföljande bakgrundsbilden, vilket är den sista bilden innan bakgrundsbilden åter ändras. En intressant del med denna bild är den tydliga skillnaden i den utkastade tomhylsans position i gråbilden (vit pil) och i eventdata (blå pil). Under dess förflyttning från vit pil till blå pil har alltså ingen visuell exponering skett, medans hela rörelsen tydligt kan ses i eventdatat.

Denna visualisering av eventdata demonstrerar hur eventkameran kan användas som en form av billig höghastighetskamera som inte genererar så speciellt hög datamängd, jämfört med vad en traditionell höghastighetskamera skulle kräva för att fånga 1500 konventionella bilder per sekund.



Figur 15 Interpolation med TimeLens-algoritmen till en bildtakt på 200 bilder/sekund jämfört med ursprungliga 25 bilder/sekund. Dessa båda interpolerade bilder ligger mellan de båda ursprungliga gråskalbilderna i Figur 13. I den vänstra bilden syns tydligt delar av mynningsflamman och i bilden till höger har rekylen börjat flytta handen med pistolen.

Det finns forskning och metoder som beskriver hur eventkameraströmmen kan användas för att interpolera informationen mellan visuella bilder. Några initiala tester med att använda ett interpolationsverktyg kallat Time Lens [9] [10] visas i Figur 15. För varje gråskalig bild har 7 interpolationsbilder genererats, vilket leder till att bildtakten ökat från

25 till 200 bilder/sekund. Hyperaktiva eventpixlar och bakgrundsbrus har filtrerats bort för att inte påverka interpolationen. Notera att det är svårt att nå en högre interpolerad bildtakt då den visuella kamerans exponeringstid måste beaktas och de interpolerade bilderna bara bär upp på tidsspannet mellan originalbildernas exponeringar. För bästa resultat bör dessutom de visuella originalbilderna innehålla så lite rörelseoskärpa som möjligt. Om temporal interpolering planeras är det därför bra med korta exponeringstider vid mätningen, vilket ger lite rörelseoskärpa och långa tidsspänn mellan originalbilderna. Det tidsspannet kan sedan fyllas med många skarpa interpolerade bilder.

I figurerna ovan är den spatiala upplösningen på det snabba förloppet (avfyrningen) hög eftersom vi filmat nära själva händelsen, vilket sällan sker i bevakningssammanhang. Upptäckt av snabba förlopp är dock av intresse även om de upptar en mindre spatial del av bilden. Tomhysan som visade i sista bilden i Figur 14 är ett sådant exempel på en snabb rörelse med relativt låg upplösning som tydligt fångas temporalt och spatialt av eventkameran.

## 4.7 Bildrekonstruktion

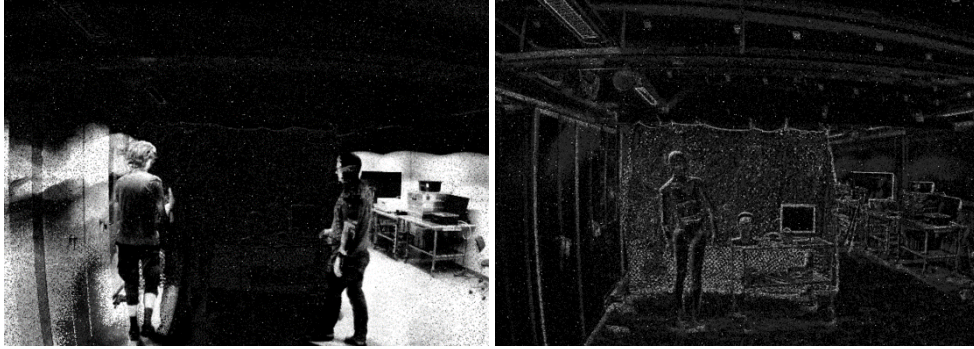
Är data från en eventkamera anonymiserande eftersom den innehåller events och inte vanliga bilder? Svaret är nej, eftersom en form av bilder kan rekonstrueras från eventdata. I Figur 16 visas två närbilder på hur ansikten kan rekonstrueras från en eventström. Gråskalenivåerna i bildrepresentationerna motsvarar inte ljusintensiteten för varje pixel, och det finns andra artefakter, men bilderna är ändå lätta för en människa att tolka. Dessa bilder har rekonstruerats med den medföljande programvaran till Inivations eventkameror. Bättre resultat är möjliga att uppnå med andra metoder som använder sig av maskininlärning, [11].



Figur 16: Detaljer av bilder från det inbyggda Accumulator-läget från Dynamic Vision Viewer [12], dvs programvaran till Inivations eventkameror. Gråskalebilder skapas utgående från inkommande events. Observera att rörliga objekt synes vara delvis genomskinliga då även events från bakgrunden kan ha påverkat pixelvärdet innan objektet kom till den positionen. Till vänster ett platt utskrivet ansikte som rörts mestadels upp och ner i bilden. Till höger en tredimensionell skyldocka som har vridits fram och tillbaka några gånger i bilden.

Några saker som kan underlätta bildrekonstruktion baserat på eventdata är följande:

- Stora objekt som rör sig och målar fram bakgrunden och sig själv, se Figur 17.
- Brus som snö och regn, eller att scenen lysas upp av en lampa med varierande intensitet, se Figur 7.
- Att kameran vibrerar kan ge ett resultat som påminner om ett kantfilter som ritar ut konturer av en stillastående bakgrund, se Figur 17.
- Att kameran rörs på andra sätt, vilket delvis diskuteras i avsnitt 4.8.



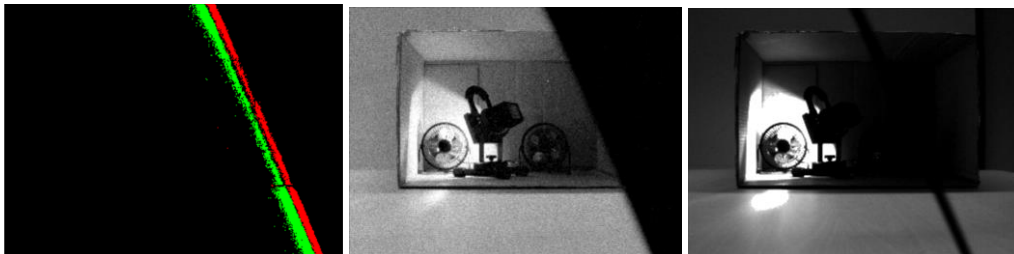
Figur 17: Bilder från det inbyggda Accumulator-läget från Dynamic Vision Viewer [12], dvs programvaran till Inivations eventkameror. Gråskalebilder skapas utgående från inkommande events. I den vänstra bilden målas bakgrunden fram efter de båda personerna som vandrar in i bilden, genom att antalet events som skapas i varje pixel och dess polaritet summeras och ger en gråskaleförändring. I den högra bilden vibrerar kameran och resultatet blir en form av kantdetektor som visar konturerna av den stillastående bakgrunden.

När ett rörligt objekt passerar en pixel skapas mer än ett event för den pixeln. Om exempelvis objektet är mörkt och ljusintaget från den statiska bakgrunden för den pixeln är relativt hög, kommer förloppet när objektet passerar pixeln bli:

1. Pixeln får bara ljus från en statisk bakgrund och inga event genereras.
2. Det rörliga objektet börjar skymma pixeln och dess ljusintag minskar vilket i snabb takt genererar att antal events med polaritet för intensitetssänkning, fram tills dess att objektet fyller hela pixelns område.
3. Objektet tar upp hela pixelns yta och om hela objektet har samma ljusstyrka genereras inga events trots dess rörelse. Om det däremot finns textur eller andra ljusförändringar i objektet kommer dessa generera events under rörelsen.
4. När så objektet avslutar passagen av pixeln kommer bakgrunden åter att börja blottas så att ljusintaget ökar och det genereras ett antal events, denna gång med polaritet för intensitetsökning, fram tills dess att objektet helt har passerat pixeln.
5. Åter igen kommer allt ljus till pixeln från den statiska bakgrunden och det generas inte längre några events.

Genom att exempelvis summera de events (och dess polaritet) som genereras i steg 4 ovan kan vi bygga upp en gråskaleförändring mellan det rörliga objektet och bakgrunden för den pixeln. Det är förenklat så som bakgrunden kan målas fram när ett objekt rör sig genom scenen.

Ifall det skulle vara önskvärt att kunna rekonstruera det som finns inom kamerans synfält när det exempelvis saknas stora rörelser eller brus finns möjligheten att montera någonting liknande ett torkarblad som sveper fram och tillbaka en bit framför objektivet. Då målas bakgrunden fram efter ett sådant svep, på bekostnad att det uppstår många events i alla pixlar, se Figur 18.



Figur 18: En statisk scen med två stillastående fläktar och en strålkastare i en låda. Scenen filmas av en DAVIS346-kamera som både samlar events och vanlig visuell bild. En svart pinne dras från vänster till höger en bit framför objektivet liknande ett torkarblad. Till vänster syns insamlade events från pinnen (rött) och det som dyker upp efter dess rörelse (grönt). I mitten syns den ackumulerade bilden som målats fram av eventen skapade av pinnens rörelse och där syns tydligt båda fläktarna. Den konventionella visuella delen av sensorn har inte använts vid skapandet av denna bild. Till höger den vanliga visuella bilden som DAVIS346 också samlar in och där är det en tydlig överexponering kring den ena fläkten medan den andra fläkten är underexponerad.

Det höga dynamiska omfånget för eventkameran gör det också möjligt att få högre dynamik i rekonstruerade bilder från eventkameraströmmen jämfört med omfånget i en vanlig visuell kamera. Detta illustreras i Figur 18 där den visuella bilden till höger båda har över- och underexponerade delar, medan den rekonstruerade bilden i mitten har ett tydligt större dynamiskt omfång.

Metoder att från en bild med låg upplösning (få antal pixlar) skapa en högupplöst bild (fler pixlar) som korrekt återger små detaljer i scenen kallas superupplösning. Ett sätt att skapa en superupplöst rekonstruktion från eventdata presenteras i [13], men har inte utvärderats för denna rapport.

Eventkameran är inte skapad för att visa en vanlig visuell bild för betraktning, men givet vissa förutsättningar kan den alltså användas för att rekonstruera olika former av semi-konventionella gråskalebilder från eventdata. Kamerans höga dynamiska omfång och snabba temporala förmåga gör det dessutom ibland möjligt att överträffa en visuell kameras möjlighet att visualisera innehållet i scenen gällande under- och överexponering samt upplösning.

## 4.8 Rörliga övervakningskameror

Många konventionella övervakningskameror har möjlighet att vrida sig i sid- och höjddled samt att zooma, t. ex. PTZ-kameror (panorering, tilt och zoom). Vissa kameror sitter istället fästa på en form av vridbord som kan vrida dem i sid- och höjddled. Zoom sker genom att kameran har ett styrbart zoom-objektiv.

Traditionellt används dessa rörelser exempelvis för:

- Svepa över omgivningen för att söka av den efter intressanta objekt som personer, fordon eller UAV:er.
- Följa intressanta objekt som exempelvis personer som rör sig.
- Zooma in på objekt för att se dem tydligare, exempelvis en persons ansikte.

Det finns algoritmer som gör detta automatiskt och väljer om de ska vara i *avsökningsläge* eller *följa upptäckta objekt-läge*. Algoritmer för sensorstyrning och –planering får då ta beslut om när det är mer intressant att söka efter nya objekt eller hur länge och ofta ett upptäckt objekt ska följas, eller när ett avsökningssvep ska avbrytas och ett tidigare sett objekt skall återfinnas.

För eventkameror tänker vi oss kanske i första hand att sensorn ska vara fast monterad för att generera få events och tydligt skilja på en stilla bakgrund och en rörlig förgrund (så som personer och fordon). Det kan dock finnas fördelar med att även kunna röra på en eventkamera. Nedan följer några olika exempel hur en rörlig eventkamera skulle kunna användas för bevakning:

- Avsökning av område för att hitta intressanta objekt. Här kan en ”step and stare” approach verka logisk där eventkameran hastigt vrids till en ny position och sedan stirrar på det området ett tag för att upptäcka rörelser (events) innan den åter snabbt vrids till en ny position. En kontinuerlig rörelse kan också vara ett möjligt alternativ men kommer då leda till en stor ström av events över troligtvis hela bilden, fast det kommer till skillnad från en konventionell kamera troligtvis inte leda till någon märkbar rörelseoskärpa.
- Följning av rörliga mål. Om ett rörligt föremål upptäcks kan eventkameran vridas för att fortsätta hålla objektet inom kamerans synfält. En fråga som bör utredas i detta fall är huruvida rörelsen ska vara av typen ”step and stare” eller en kontinuerlig följdning.
- Bildrekonstruktion från eventkameraströmmen. Om ingen konventionell kamera finns tillgänglig kan rörelse av eventkamera användas till bildrekonstruktion för att skapa en mer visuell bild av området liknande beskrivningen i föregående sektion. Dock behövs komensation tas för själva rörelsen då position på objekten

och bakgrunden i bilden flyttas till nya pixlar genom rörelsen, [11]. Sådan bildrekonstruktion kan exempelvis användas för att se vilken typ av objekt som en person ställde på marken, eller bekräfta att en bil fortfarande står parkerad mitt i bilden.

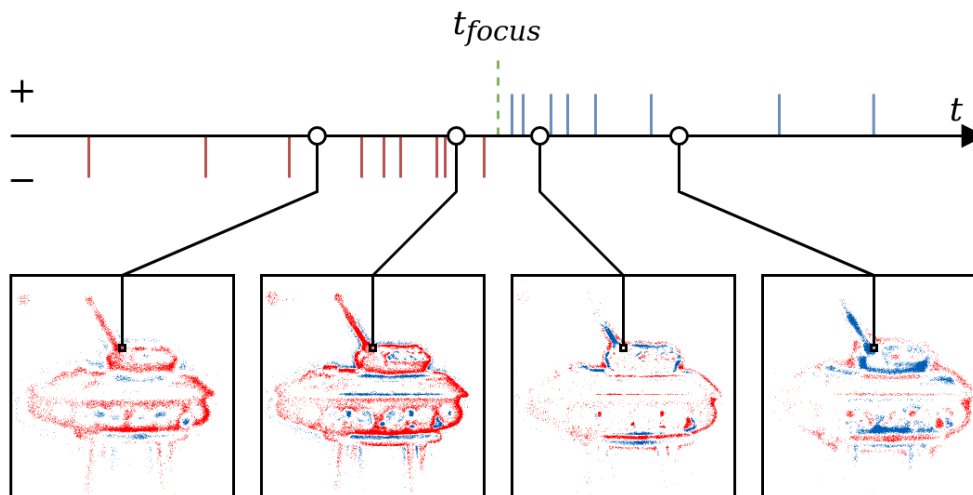
- Med ett zoomobjektiv på en eventkamera kan vid inzoomning högre upplösning nås vid bildrekonstruktion, små objekt kan ges fler event och det kan bli enklare att exempelvis avläsa frekvenser på propellrar eller andra detaljer på fordon. En in- eller utzoomning kommer dessutom troligtvis kunna användas för bildrekonstruktion, enligt föregående punkt.

Algoritmer för automatisk sensorstyrning och -planering kommer behöva ta hänsyn till eventkamerans unika egenskaper och inte använda samma approach som för en konventionell kamera, eftersom det är stor skillnad på eventströmmen mellan en rörlig och stillastående kamera.

## 4.9 Djupinformation genom fokussvep

En intressant parameter som går att vrida på eventkameran är dess fokus. Det går nämligen få fram djupinformation om scenen genom att utföra ett så kallat fokussvep, från exempelvis fokus nära sensorn bort till oändligheten. Vi har undersökt den metod som presenterar i [14].

Denna djupinformation gör att det vid detektion av intressanta objekt snabbt går att styra kamerans fokus mot dessa och avbilda dem skarpare, vilket i sin tur t.ex. kan underlätta fortsatt informationsinhämtning såsom objektsklassificering och identifiering. För stillastående eller långsamma objekt kan denna djupinformation dessutom skapa en enklare typ av 3D-information av objektet som i förlängningen kan vara användbar för automatisk klassning, men även för exempelvis forensiska undersökningar där ofta storlek och avstånd mellan detaljer är av intresse.

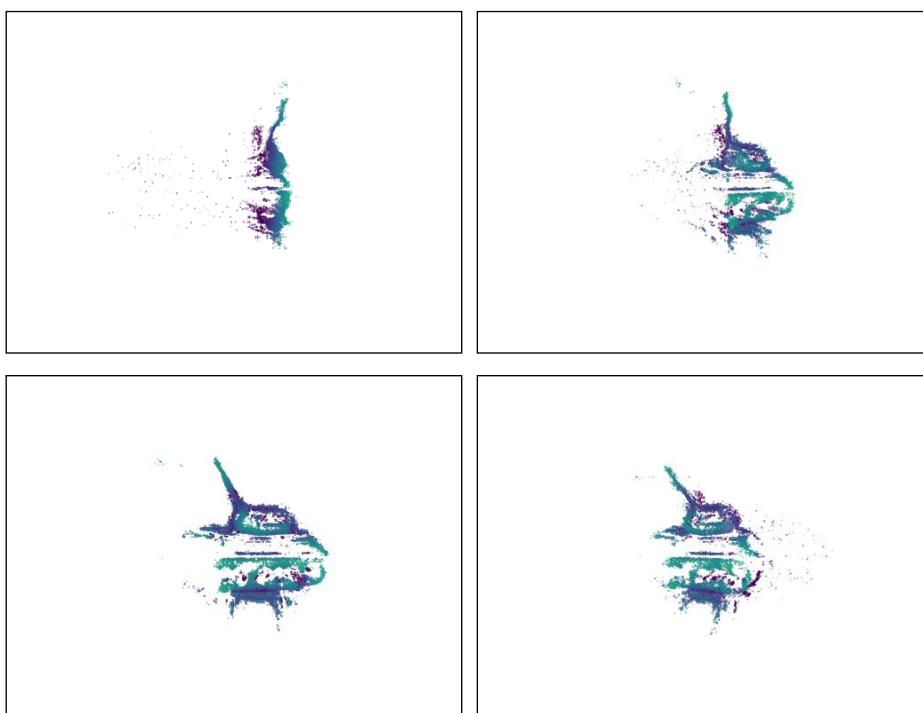


Figur 19 Illustration på den signal som en individuell pixel i eventkameran genererar under ett fokussvep, över tiden  $t$ . Scenen består av en modell av en stridsvagn placerad på en pall. När fokusplanet rör sig mot objektet genereras eventen på den pixeln med en bestämd polaritet (i det här fallet negativa - röda). Från tidpunkten att fokusplanet når objektet är eventen av inverterad polaritet (positiva - blåa). Genom att hitta när ( $t_{focus}$ ) under fokussvepet inversionen (polaritetskitet) sker kan fokus för den punkten hittas, och därmed djupinformation erhållas.

I Figur 19 visas ett exempel på ett fokussvep där scenen vi mätt upp består av en stationär modell av en stridsvagn som står på en pall. I detta fall användes ett mekaniskt objektivet som styrs elektroniskt för att skapa fokussvepet. Figuren illustrerar vad som händer pixeln vid kanons fäste i tornet, med en tidslinje för dess events, samt fyra exempelbilder på events i bilden ackumulerade under några kortare tidsintervall. Under första halvan av

fokussvepet får vi events med negativ polaritet i pixeln, med ökande frekvens. Vid en tidpunkt  $t_{focus}$  ändras polariteten och därefter uppstår positiva events under återstoden av svepet, med fallande frekvens. Tidpunkten  $t_{focus}$  inträffar då pixeln är i fokus.

På liknande sätt kan under samma svep övriga pixlars fokus i sensorn undersökas. Dock uppstår bara tydliga polaritetsinversioner för de pixlar som ligger vid objektets konturer, texturer och andra visuella kanter. Övriga pixlar genererar troligtvis också events under fokussvepet men utan den karaktäristiska ökande frekvensen av events, följt av ett snabbt polaritetsskifte och sedan minskande frekvens, som uppstår när fokusplanet passerar kanter. Därmed kan ointressanta events filtreras bort så att bara intressanta pixlar erhåller ett djupestimater.



Figur 20 Fyra vyer på ett punktmoln  $(x, y, z)$  för djupet i ett antal pixlar taget från fokussvepet. Varje punkt motsvarar en pixel som uppvisat ett godkänt polaritetsskifte. Grönare punkter ligger närmare kameran. Modellen av stridsvagnen och övre delen av pallen som den står på kan urskiljas och det är också tydligt att dess kanon är riktad lite snett mot kameran.

Ett sätt att illustrera uppmätt djupestimater är att plotta pixlarna  $(x, y)$  med en tredje dimension där tidpunkt  $t_{focus}$  ersatts med djupet  $z$ . I Figur 20 visas en sådan 3D-rendering av djupskattningen från fyra olika perspektiv. Kanontornet hos stridsvagnsmodellen var under mätningen något vinklad mot kameran vilket också fångades i djupestimater. Detta är särskilt tydligt i första perspektivet sedd framifrån stridsvagnen. I andra perspektiv framstår kanontornets profil och delar av fordonskroppen tydligt i denna avbildning utav djupestimater.

## 4.10 Strömsnål övervakning

Vid automatisk övervakning av exempelvis avlägsna platser kan låg effektförbrukning vara önskvärd för att kunna driva systemet på batteri. En fast monterad eventkamera bör, överlag, generera få events och därmed förbruka låg effekt. Även effekten som behövs för att driva de automatiska algoritmerna måste dock räknas med i systemet. Där finns ansatser med de tidigare nämnda spikande neurala nätverken som är snabba och ger låg effektförbrukning. I kombination med neuromorfisk beräkningshårdvara (se avsnitt 3.3)

finns potential för system som utför relativt kvalificerade uppgifter med en energibudget på tiotals milliwatt.

En förevändning mot att använda eventkameror skulle kunna vara att eventdata är svårtolkat av en operatör, ibland även efter bildrekonstruktion. I fall där en algoritm upptäcker störningar och hot behöver dock inte alltid data vara helt anpassade för en människa, vilket gör att man kan komma undan med billigare, energisnålare sensorer och snabbare system utan att för den sakens skull tumma på övervakningsförmågan.

Om visuella bilder trots allt krävs skulle eventkameran kunna styra så att en stillbild tas med en strömsnål konventionell kamera som exempelvis sitter bredvid eventkameran, lite som funktionen i en åtelkamera. Alternativt kan en mer inzoomad visuell kamera styras in mot den del av scenen där eventkameran upptäckt någonting intressant.

Under 2024 planeras strömsnåla eventkameror och beräkningshårdvara att undersökas närmare.

## 5 Analys och slutsatser

I denna rapport har både sensorers egenskaper, samt möjligheter för användning inom bevakning, presenterats. Detta kapitel innehåller en bedömning av TRL-nivå och översiktliga slutsatser och värdering av eventkameran. Vidare diskuteras några nya möjligheter som kan uppnås med eventkameran och som är svåra eller omöjliga att uppnå med konventionella övervakningskameror.

### 5.1 TRL-nivå

För en bedömning av eventkamerors tekniska mognad och TRL-nivå<sup>4</sup> är det lämpligt att göra en uppdelning i följande delar:

- 1) Sensorers mognad.
- 2) Signalbehandlingens mognad (här tillkommer mognaden hos hårdvara som stöder signalbehandlingen).
- 3) Tillämpningar inom bevakning.

Eventkameran och den bakomliggande tekniken är idag relativt mogen och dess TRL-nivå bedöms vara 6-7. Eventkameran är en hyllvara som kan beställas för omgående leverans. Installation för att komma igång tar någon minut och sedan har man en fullgod eventkamera för utvärdering och enklare försök. De sensorer som FOI testat uppfyller i stort sett leverantörernas prestandaspecifikationer.

Signalbehandlingen på eventdata ligger mognadsmässigt efter i jämförelse med sensorutvecklingen och dess TRL-nivå bedöms ligga på 4-5. Signalbehandlingen av eventdata är dock ett aktivt forskningsområde där mycket har hänt under relativt kort tid. En ansats som frekvent förekommer inom signalbehandling är att man försöker ompaketera eventdata till någon annat format – t.ex. bilder – och sedan använda tidigare utvecklade metoder på dessa. Man har bland annat använt maskininlärning för detektion och klassificering av bilder genererade från eventdata med goda resultat. Problemet med denna typ av ansatser är att man inte kan utnyttja sensorers unika egenskaper. Hårdvara för energieffektiv signalbehandling, vanligen med maskininlärning som spikande neurala nätverk (SNN), på eventdata finns. De är fortfarande på en låg TRL-nivå, omkring 3-4.

Erfarenheter från de få konkreta tillämpningarna som finns utanför forskning och utveckling indikerar att hela sensor kedjan för eventkamera ännu inte är tillräckligt mogen för tillämpningar inom bevakning eller andra områden som tillverkningsindustri och robotik. Tillverkarna av eventkameror tror mycket på sina produkters framtida användning i bilar och inom robotik, men bolag inom dessa marknader av ännu inte börjat använda sensorer. Vi bedömer att nuvarande TRL-nivå på tillämpningar ligger på 2-4, men att den kommer öka hastigt när utvecklingen av signalbehandlingen kommit längre. Sammanfattat beskrivs mognaden för införande av eventkameran i Tabell 2

Tabell 2: TRL-nivå på olika delar av eventkamerans tekniska mognad.

Mognad för eventkameran	TRL-nivå
Sensorn	6-7
Signalbehandling	4-5
Signalbehandling - SNN	3-4
Tillämpningar inom bevakning	2-4

<sup>4</sup> TRL - Technology readiness levels. En niogradig skala (1-9) på hur mogen en teknologi är. Från grundforskning (1) till system beprövat i operativ miljö (9). Vi utökar i detta kapitel begreppet något genom att även inkludera mognad av tillämpning av teknologi, och sätter in den i samma skala.



## 5.2 Värdering

Eventkameran har flera unika egenskaper som potentiellt ger nya möjligheter. Framför allt är det den temporala upplösningen som öppnar upp för flera nya förmågor. Den höga temporala upplösningen gör att man direkt kan upptäcka snabba händelser, såsom explosioner eller mynningsflammor, och även lösa upp dessa förlopp tidsmässigt. Eventkameran möjliggör också upptäckt och analys av högfrekventa processer så som blinkande ljus, propellrar eller högfrekvent vibration. Frekvensanalys är ett väldigt starkt verktyg som kan användas för att upptäcka och klassificera objekt, men också för att filtrera bort bakgrundsklotter så som blinkande ljus eller vajande grenar. Konventionella kameror saknar en stor del av denna förmåga och arbetar sällan på högre uppdateringshastighet än 50 bilder/sekund, och andra sensorer som kan uppfatta snabbare förlopp så som radar, saknar den spatiala information som en eventkamera har.

Eventkamerans höga dynamiska omfång gör att den till skillnad mot konventionella kameror kan arbeta på både mörka och ljusa delar av scenen på samma gång och det minskar risken att objekt och händelser döljs av över- eller underexponering. Detta ökar sannolikheten att kunna upptäcka vad som händer i hela scenen och därmed få en bättre lägesbild, även vid svåra förhållanden som kombination av solljus och skuggor samt vid lågljuszförhållanden.

Låg bandbredd och låg strömförbrukning är två egenskaper som också karakteriserar eventkameran. Sensorn genererar bara event när något förändras i scenen och sensordata analyseras enbart när event genereras. I statiska scener händer väldigt lite vilket resulterar i låg bandbredd. Vidare kan strömförbrukningen för både sensorn och beräkningsenheten hållas låg då få events genereras. För statiska scener med begränsad aktivitet är den potentiella drifttiden för batteridrivna enhet relativt lång. I vissa tillämpningar är det troligt att eventkameran kommer användas för invisning av andra mer effektkrävande sensorer, som bildalstrande högupplösta kameror eller aktiva sensorer som radar, för att få kompletterande data om scenen.

Eventkamera kombinerat med en välanpassad beräkningsenhet har stor potential inom bevakning. Införande av tekniken hindras dock av begränsningar inom signalbehandlingen där metoder som är exekverbara på energieffektiv hårdvara och som drar nytta av sensorns unika egenskaper fortfarande håller på att utvecklas. Eventkameror med god prestanda till överkomliga priser är idag hyllvaror och det är inte längre begränsningar hos sensorn som hindrar införandet av tekniken på bred front. Forskningen kring signalbehandlingen är mycket aktiv och nya lovande resultat presenteras kontinuerligt. I och med den ökade tillgången på sensorer och den höga aktiviteten inom forskningen kommer begränsningar inom signalbehandlingen att minska. När kommersiella bevakningslösningar kommer att erbjudas är i dagsläget svårt att avgöra.

Konventionella kameror har i olika former funnits i över hundra år och är i första hand skapade för att visa bilder för människor, medan eventkameror har unika egenskaper som inte behöver begränsa sig till vad vårt synsinne kan uppfatta utan lämpar sig för automatiska signalbehandlingsalgoritmer och att generera ny information för en bättre situationsuppfattning.

## 6 Referenser

- [1] Jens Grundmark, Hanna Hamrell, Johan Berneland & David Gustafsson, "Frequency Analysis on Neuromorphic Data Using Discrete Fourier Transform," i *IPTA 2023*, Paris, 2023.
- [2] Henri Rebecq, Daniel Gehrig & Davide Scaramuzza, "ESIM: an Open Event Camera Simulator," i *Proceedings of The 2nd Conference on Robot Learning*, 2018.
- [3] SET-ET-131, "Applications of Neuromorphic Cameras," 2023. [Online]. Available: <https://www.sto.nato.int/Lists/test1/activitydetails.aspx?ID=17240>.
- [4] David Bergström, David Gustafsson, Sebastian Möller & Erik Valldor, "Eventbaserad avbildning - En orientering om tekniken och dess potential," FOI Memo 7660, 2021.
- [5] Hanna Hamrell, Johan Berneland, Jens Grundmark, Oskar Karlsson, Hannes Ovrén & David Gustafsson, "Neuromorphic Cameras - An Introduction with Initial Results," i *SSBA 2023*, Kolmården, 2023.
- [6] Ignacio Alzugaray & Margarita Chli, "Asynchronous corner detection and tracking for event cameras in real time," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, p. 3177–3184, Oct 2018.
- [7] Kashu Yamazaki, Viet-Khoa Vo-Ho, Darshan Bulsara & Ngan Le, "Spiking neural networks and their applications: A review," *Brain Sci.*, vol. 12, no. 7, p. 863, Jun 2022.
- [8] Guillermo Gallego, et al, "Event-Based Vision: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 154-180, 2022.
- [9] Stepan Tulyakov, Daniel Gehrig, Stamatios Georgoulis, Julius Erbach, Mathias Gehrig, Yuanyou Li & Davide Scaramuzza, "Time Lens: Event-based Video Frame Interpolation," i *IEEE Conference on Computer Vision and Pattern Recognition*, 2021.
- [10] Stepan Tulyakov, Alfredo Bochicchio, Daniel Gehrig, Stamatios Georgoulis, Yuanyou Li & Davide Scaramuzza, "Time Lens++: Event-based Frame Interpolation with Parametric Non-linear Flow and Multi-scale Fusion," i *IEEE Conference of Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 2022.
- [11] Henri Rebecq, René Ranftl, Vladlen Koltun & Davide Scaramuzza, "High Speed and High Dynamic Range Video with an Event Camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.
- [12] IniVation, 2023. [Online]. Available: <https://inivation.com/developer/software/>.
- [13] Lin Wang, Tae-Kyun Kim & Kuk-Jin Yoon, "EventSR: From Asynchronous Events to Image Reconstruction, Restoration, and Super-Resolution via End-to-End Adversarial Learning," i *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
- [14] Germain Haessig, Xavier Berthelon, Sio-Hoi Ieng & Ryad Benosman, "A Spiking Neural Network Model of Depth from Defocus for Event-based Neuromorphic Vision," *Scientific Reports*, 2019.



## 7 Bilaga: Händelser som önskas upptäckas vid bevakning

Vid bevakning av en anläggning, även enskilt belägna byggnader, finns det händelser som bör upptäckas av ett sensorsystem. Några av dem beskrivs nedan.

1. Närma sig anläggningen
  - a. Personer går, springer mot anläggningen
  - b. Kör markfordon mot anläggningen: bil, lastbil, UGV, grävmaskin etc.
  - c. UAV flyger mot anläggningen
2. Bredvid anläggningen
  - a. Personer: går runt, sitter still länge, uppehåller sig/dräller (loitering), återkommer ofta, är där i mörker, lämnar av något, tar upp något, klättrar över staket
  - b. Markfordon: kör runt anläggningen, parkerar, återkommer ofta, är där i mörker, lämnar av något, tar upp något
  - c. UAV: flyger runt, hovrar, släpper last
3. Placerar något (mindre objekt/sak) bredvid eller på anläggningen
  - a. Explosivämne
  - b. Verktyg
  - c. Kamera
  - d. Annan elektronisk utrustning
4. Skadar anläggningen
  - a. Klottrar
  - b. Bryter upp dörr, fönster, vägg, tak
  - c. Svetsar igen dörr eller fönster, spikar igen fönster
  - d. Utlöser sprängladdning
  - e. Skadar kamera och andra övervakningssystem/utrustning för att bevaka anläggningen
  - f. Skadar staket eller grind
5. Kartlägga anläggningen
  - a. Person/personer, markfordon, UAV återkommer ofta
  - b. Fotografera/filma med RPAS, person på plats eller utplacerad kamera
  - c. Placera utrustning för att mäta elektroniska och magnetiska signaler
6. Stjäla från anläggningen
  - a. Stjäla kameror och annan övervakningsutrustning
  - b. Stjäla elkablar, dörrar, fönster, metallbleck etc.
  - c. Stjäla det som förvaras i anläggningen

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](http://www.foi.se)