

Heterogen Sensorfusion: Verifikation av Tracking

Henrik Lindelöf Bilski
Lunds Tekniska Högskola, Lund

Introduktion

Tracking är det matematiska begreppet att spåra objekt över tid. Det kommer från radarforskningen under andra världskriget och är idag en viktig teknik som är brett använd. Det går och är ofta fördelaktigt att kombinera flera radarsensorer, vilket kallas sensorfusion. Med fler sensorer går det att täcka större yta, men framförallt att spåra objekt med fler än en sensor samtidigt. Då varje sensor har ett feltal som inte är beroende av dem andra sensorerna går det att isolera och minimera feltalen om man har tillräckligt många sensorer.

Även om radar är vanligast för tracking kan man utföra det med en kamera. Med rätt matematik går det att följa personer som rör sig i bilden. Till skillnad från i radar så spåras positionerna i kameran i bildplanet, vilket inte är direkt översättbart till verklig position, då kameror har perspektivfel.

Eftersom fler sensorer ger en klarare bild, borde även flera olika sensorer ge en klarare bild. Detta medför svårigheter då kamera och radar inte ser objekt på samma sätt. Systemet kallas heterogent, eftersom en eller fler dimensioner skiljer sig mellan sensorerna. Radar mäter avstånd och infallsvinkel. Kameran kan spåra i skärmkoordinater. Infallsvinkeln och den horisontella skärmkoordinaten är i samma dimension, men med två olika mått. Vertikal skärmkoordinat och avståndet är två helt skilda dimensioner, och kan inte översättas till varandra med någon pålitlighet.

Frågeställningen är därför att kombinera dem heterogena sensorerna kamera och radar i ändamålet att spåra saker. Med en testmontering med både kamera och radar, är det möjligt att förbättra tracking genom att sammanfoga sensorernas data?

Metodik

Trackingen beräknas med **Probabilistiskt Data Associations Filter** (PDAF). Det är en vidareutveckling av ett enklare filter, Kalman filtret. Kalman filtret är ett optimalt filter för en viss typ av felminimering. PDAF är en förlängning av Kalman filtret då man har flera objekt och träffar i en sensor. Det är ett associationsproblem att associera rätt träffar med rätt objekt. PDAF matchar varje objekt med träffar som ligger närmast i sannolikhetsplanet.

PDAF kan förlängas till fall med flera sensorer, men det finns andra metoder specifikt gjorda för sensorfusion. Fördelen med PDAF är att det är enkelt och snabbt. Nackdelen är att den behövs anpassas för flera sensorer och antalet parametrar växer snabbt. Antalet parametrar blir över 30 stycken om vi räknar delar av matriser separat.

PDAF och Kalman filtret baseras på gör en uppskattning över var objektet kommer att befinna sig, sedan jämför den uppskattade positionen med omkringliggande data. Ifall det finns data blir trackingen säkrare på objektet samt justerar position och riktning efter avvikelser. Saknas det data blir trackingen mindre säker på positionen

men antar att uppskattningen stämmer, då det händer att sensorer missar objekt tillfälligt. Saknas data för ett objekt för länge slängs objektet och antas vara försvunnet. Det är antaget att datan kommer till trackingen regelbundet och att avståndet mellan sensorerna är tillräckligt lågt för att marginaliseras bort.

Sensorfusion

För att sammanfoga flera sensorer behöver deras data se liknande ut. Kamera informationen i skärmkoordinater behöver översättas till vinklar för att kunna jämföras med radardatan. När väl datan har samma form, går det att matcha radardata och kameradata. Då systemet är heterogent går det inte att matcha alla dimensioner, utan bara infallsvinkeln är en delad dimension i båda datamängderna.

När uppskattningen sker i PDAF och det finns flera sensorer görs en uppskattning för varje sensor. Två olika lösningar för att anpassa uppskattningen till datan testades, sekvensiell och parallell. I sekvensiell fusion sker uppdateringen först med ena sensorn, och sedan den nya uppskattade positionen anpassas med data från den andra sensorn. I parallell fusion sker två separata anpassningar, som sedan kombineras genom en fusionsregel till en gemensam position.

Utvärdering

För att objektivt kunna jämföra de olika metoderna behövs ett oberoende facit. För att göra ett facit behövs tester. Över en timme av film med både kamera och radar spelades in, med varierande längd, miljö, och objekt. Facit konstruerades med ett användarinterface och manuell inmatning. Att använda manuellt skapande av facit har lägre precision och tillförlitlighet än att fysiskt mäta ut facit, men det görs mycket snabbare och med mindre utrustning.

Med facit och tester gäller det att utvärdera metoderna mot facit. Bedömningen valdes att göras med Clear MOT Metrics [1]. Clear MOT är tidsinvariant då den bedömer utifrån frekvensen av fel, inte hur eller när. Bedömningen är indelad i två delar, en precisionsdel och en träffsäkerhetsdel. Här är vi mest intresserade av träffsäkerhet då precision är otydligt definierad när antalet dimensioner varierar. Träffsäkerhet har samma betydelse oavsett antalet dimensioner, eftersom antingen hittas ett objekt eller så hittas det inte.

$$MOTA = 1 - \frac{\sum f_p + \sum f_n + \sum id^-}{\sum \forall p} \quad (1)$$

Ekvation (1) är fulla ekvationen för att bedöma träffsäkerhet enligt Clear MOT. $\sum f_p$ är antalet falska positiva, tillfällen då trackingen säger att den har hittat något fast inget objekt finns där. $\sum f_n$ är antalet falska negativa, tillfällen då trackingen misslyckas att hitta ett objekt som finns i facit. $\sum id^-$ är antalet gånger ett objekt som ska hittas får fel id, vilket händer då trackingen blandar ihop två objekt eller mitt under trackingen av ett objekt tar bort det och skapar ett nytt. $\sum \forall p$ är antalet sanna objekt i facit.

En till viktig punkt för objektiv bedömning är att det inte är implementationsfel i metoderna. Detta inkluderar också parametrarna som metoden förlitar sig på. Det gäller att välja bra parametrar för att metoderna ska fungera väl. Det ursprungliga sättet att välja parametrar var genom testning. Parametrar valdes, resultaten spelades upp, och parametrarna korrigerades. Detta var en långsam och tidsödande process.

För att slippa att manuellt anpassa parametrar valdes det att sköta det automatiskt. Eftersom det finns ett facit, en metod, och en bedömning går det att matematiskt optimera parametrarna så att metoden blir så bra som möjligt. Optimeringsmetoden som valdes heter **Steepest Descent**. Steepest Descent letar efter maximum, toppar i rymden av alla värden. Det finns ingen garanti på att Steepest Descent hittar den högsta toppen/globalt maximum, men den kommer vandra upp för någon topp. Även här gick förbättringen av parametrar långsamt, men med fördelen att det kan ske i bakgrunden utan arbete från användare.

Resultat

Resultat för dem olika metoderna presenteras nedan. Fusion är bättre på att spåra objekt genom att kombinera kameradata och radardata, även om dem är heterogena.

Metod	Träffsäkerhet
Sekvensiell Fusion	54.2%
Parallell Fusion	58.2%
Endast Radar	29.8%
Endast Kamera	50.1%



Figure 1: Visualisering av tracking. Vänstra siffran är id och högra siffran är avstånd från sensorn i meter.

Sammanfattning

- Heterogen Sensorfusion är bättre på trackingen än sensorerna separat
- PDAF räcker för att utföra heterogen sensorfusion
- Antalet parametrar i PDAF är ett problem, men kan lösas matematiskt

Referenser

- [1] K. Bernardin and R. Stiefelhagen, Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics, *EURASIP Journal on Image and Video Processing*, 2008.