

# Kamerabaserade tracking-tekniker för mobil Augmented Reality

## Inledning

Denna rapport redovisar ett fördjupningsarbete inom kursen ”Att använda Virtual Reality: nya möjligheter inom olika tillämpningsområden”, som är en kurs på 7,5 hp vid Lunds tekniska högskola.

Anledningen till att jag valt Augmented Reality som fördjupningsämne är att jag tycker det är ett mycket intressant område, samt att det även är ett högaktuellt ämne, eftersom tekniken som används för mobil Augmented Reality nu blivit så pass bra och utrustningen som krävs så pass liten, att det börjar komma intressanta tillämpningar på marknaden.

Vissa ord i titeln ovan och en del av uttrycken i denna text är skriven på engelska eftersom det ibland är svårt att översätta engelska tekniska begrepp till bra svenska. Vid några tillfällen har jag skrivit både det svenska och engelska begreppet för att förtydliga. Titeln lyder på engelska ”Vision-based tracking techniques for Mobile Augmented Reality” och ett försök att skriva titeln helt på svenska skulle kunna vara ”Kamerabaserade spårningstekniker för mobil förstärkt verklighet”, men med tanke på detta sammanhang och målgrupp anser jag att det blir mer begripligt att behålla vissa, mer eller mindre, vedertagna uttryck på engelska. Eventuellt kan ”kamera” ersättas med ordet ”bild”, eller bytas mot det mer generella uttrycket ”[datorseende](#)” (”[computer vision](#)”) [42, 43].

Ett tillägg till titeln skulle kunna vara ”I monokulära system; en kamera”, eftersom jag har inriktat mig på att betrakta system där endast en (1) kamera används, till skillnad från system där två kameror används för att skapa ett djupseende.

I rapporten kommer jag att presentera vad Augmented Reality är, exempel på användningsområden, hårdvara/utrustning för mobil Augmented Reality, samt olika ”tracking”-tekniker med fokus på kamera-/bildbaserade tekniker och metoder.

Fakta och underlag i form av vetenskapliga artiklar har jag sökt upp i huvudsak via [IEEE Xplore Digital Library](#) [26] och [ACM Digital Library](#) [16]. De flesta av dessa artiklar är presenterade vid den årligt återkommande konferensen [ISMAR](#) (International Symposium on Mixed and Augmented Reality) [28]. I övrigt har jag funnit matnyttiga uppgifter på bl a sajterna [Games Alfresco](#) [20, 21] och [UgoTrade](#) [39]. Tidningsartiklar samt webbsidor tillhörande universitet och företag som forskar och utvecklar inom området är ytterligare informationskällor. Dessutom har jag utnyttjat YouTube för att söka fram intressanta illustrerande exempel, och Wikipedia för faktauppgifter. En referenslista finns i slutet av detta dokument.

## Augmented Reality

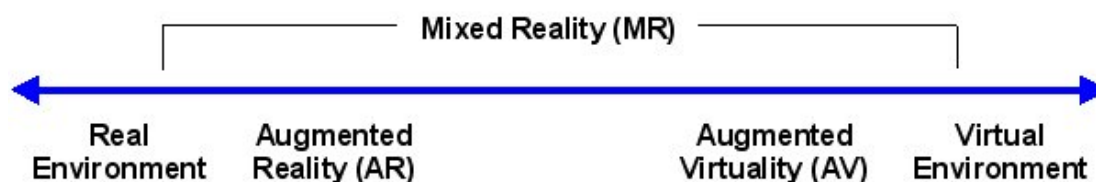
Augmented Reality – förstärkt verklighet – innebär att man kombinerar reell och virtuell verklighet (datorgenererade bilder) som kan betraktas genom:

- ”see-through Head Mounted Displays” (HMD)
- projektions-baserade / spatiaa displayer
- handhållna displayer (tablet PC, smartphone) [9, 41]



*Figur 1. En ”tablet PC” med en Augmented Reality-applikation för hjälp vid reparation av en skrivare [9].*

Skillnaden mellan Virtual Reality (VR) och Augmented Reality (AR) är att i VR är användaren totalt omgiven av den virtuella världen, medan i AR förstärks den reella verkligheten med digitala objekt. Ett förslag på att beskriva förhållandet mellan AR och VR kan ses i figur 2.



Figur 2. *Reality-Virtuality Continuum* [15, 41, 50].

Syftet med AR är att förstärka verkligheten genom att ge användaren information, upplevelser och guidning i den omgivning som användaren befinner sig i. Förstärkningen är i realtid och kontextberoende [41].

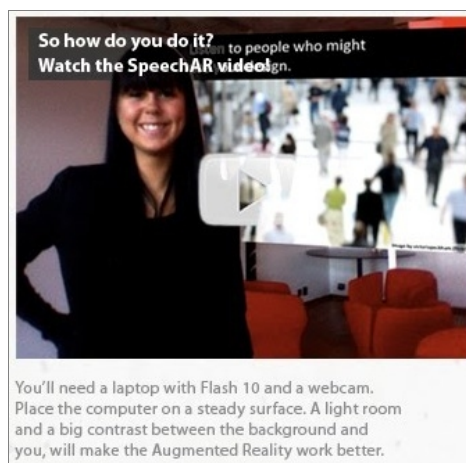
En allmänt vedertagen definition på Augmented Reality är Ronald Azumas definition från 1997 [2, 3, 9, 15]:

- kombinerar den reella verkligheten med virtuella objekt.
- är interaktiv i realtid (samt är kontextberoende).
- återges i 3D ( och är rumsligt ”förankrad”).

Anledningen till att jag i den andra punkten lagt till ordet kontextberoende, och lagt till uttrycket ”rumsligt förankrad” i tredje punkten, är att begreppet Augmented Reality på senare tid blivit något urvattnat, eftersom det blivit ett populärt begrepp som kommit att användas i sammanhang som egentligen inte handlar om Augmented Reality, om man ska rätta sig efter Azumas definition.

Som exempel på denna urvattning av begreppet AR kan jag nämna en artikel i Metro Teknik [10], där en entreprenör vid namn Michaela Schmied påstår sig använda Augmented Reality-teknik i applikationen [SpeechAR](#) [37] som hon utvecklat (se figur 3). Applikationen är ämnad för presentationsteknikträning och går ut på att man filmar sig själv med en webb-kamera och denna bild återges på din datorskärm, tillsammans med en PowerPoint-presentation som visas i ett hörn av bilden.

Om vi tittar på Azumas definition av Augmented Reality så kan vi kanske kalla PowerPoint-bilden för ett virtuellt objekt som kombineras med bilden från webb-kameran, men när vi kommer till punkt 2 och 3 passar applikationen inte längre in i begreppet Augmented Reality. PowerPoint-bilden är inte på något sätt interaktiv med omgivningen, och hur du än riktar kameran – upp i taket eller ned på golvet – kommer bilden fortfarande bara att ”hänga kvar” i samma hörn. Dessutom återges grafiken inte i 3D (även om detta krav är uppluckrat i de flesta befintliga applikationer för smartphones, se nedan) och grafiken är inte på minsta sätt placerad eller förankrad i rummet. Det är inget fel på applikationen i sig, men att kalla det för Augmented Reality är lite magstarkt. Det handlar istället helt enkelt om en [bild-i-bild-funktion](#) [47], en teknik för TV-monitorer, och som varit möjlig sedan 70-talet.



Figur 3. *SpeechAR* [37].

## Användningsområden

Det behöver inte vara fotorealism på 3D-grafiken i Augmented Reality. Enkla AR-applikationer kan t ex bestå av pilar för att visa riktningen, mer avancerade applikationer kan generera bilder som knappt går att skilja från verkligheten. Avancerade applikationer finns t ex inom medicinområdet, eller inom byggområdet, där ”röntgenbilder” kan förstärka verkligheten för att underlätta arbetet [9].

Andra områden kan vara hjälp att utföra uppgifter, t ex en reparation, eller för militära ändamål, för underhållning, etc. AR kan också användas för att för att visualisera inom arkitektur, inredning och arkeologi. Se gärna ”[Augmented Reality](#)” på Wikipedia [41] för fler exempel på användningsområden.

Populära befintliga program för smartphones (iPhone m fl) baseras på t ex [Layar](#) [34], som dock inte använder någon teknik för bildanalys, utan nyttjar enbart sensorer som GPS och kompass för att fastställa position och riktning. Ett exempel är företaget [GoWeb3D](#) [23] som byggt [applikationer i Layer](#).

[TAT](#) (The Astonishing Tribe) har utvecklat den uppseendeväckande prototypen ”[TAT augmented ID](#)” [38] (se figur 5 och tillhörande video), som snart sägs bli en färdig applikation.

Dock är den förstärkta verkligheten i dessa mobila applikationer i de flesta fall (än så länge) inte i 3D, utan vanlig 2D-text. Applikationer med 3D-grafik är i antågande, exempelvis baserat på [Layar3D Augmented Reality Engine](#) [35].

Frågan är om det inte i vissa av dessa [applikationer som t ex GoWeb3D utvecklat](#), hade fungerat lika bra, eller kanske t o m bättre, med enbart en 2D-karta där objekten hade markerats på kartan, istället för 2D-text som visas ”hängande i luften”.

## Utrustning/hårdvara för mobil AR

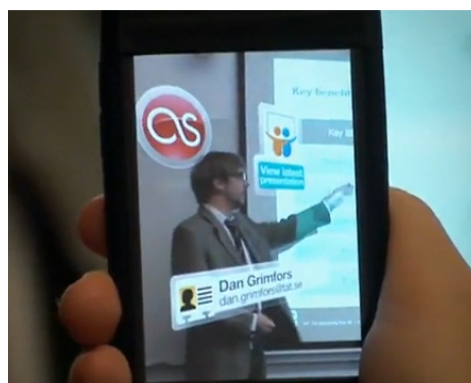
Redan 1968 skapade Donald Sutherland en HMD för Virtual Reality [27], men begreppet Augmented Reality myntades först 1992 av Tom Caudell och David Mizell [27, 41].

Som jag nämnde tidigare börjar nu hårdvaran som krävs för mobil AR bli så pass kraftfull i förhållande till storleken, samtidigt som den blir billigare och mer spridd, att mobil AR verkligen är något som är på frammarsch i vardagen. Inte minst att dagstidningen Sydsvenskan nyligen ägnade nästan ett helt uppslag åt [ett reportage om ”förhöjd verklighet”](#) (AR) [5], visar på ett begynnande mer allmänt intresse.

Utvecklingen för hårdvara för mobil AR har gått från tung och otymplig utrustning placerad i en ryggsäck, till nutidens smartphones, och vidare till framtidsvisioner om kontaktlinser.



*Figur 4. [Ghostwire](#) [22] är ett svensktutvecklat prisbelönt AR-spel för Nintendo DSi. Mjukvaran analyserar videobilden i realtid [1].*



*Figur 5. [TAT augmented ID](#) [38], baseras på mjukvara för ansiktsgenkänning från [Polar Rose](#).*

Här följer en kort historik respektive framtidsvision över utrustning för mobil AR [20]:



*Figur 6. Generation "Kit Bag".*

Generation "Kit bag",

för ca 10 år sedan och längre tillbaka:

- ryggsäck med bärbar dator samt "Head Mounted Display" (HMD)
- användes enbart inom forskning
- tung, komplex och dyr utrustning



*Figur 7. Generation "Hand Bag".*

Generation "Hand bag",

från 2003 till nutid:

- skärmarna på mobiler blir större och hårdvaran kraftfullare
- massproduktion av lämplig utrustning – prismässigt överkomligt
- lätta att bära men händerna är upptagna
- ger en begränsad upplevelse



*Figur 8. Generation "No Bag".*

Generation "No bag",

kommande 3–15 åren:

- glasögon och därefter kontaktlinser
- "handsfree"
- ger en större helhetsupplevelse

## Tracking-tekniker

När man fördjupar sig inom olika tracking-tekniker kommer man snart in på algoritmer och matematiska modeller. Jag har dock inte här fördjupat mig i tracking på matematisk nivå.

Man kan dela in tracking-tekniker i tre kategorier [3]:

- Sensor-baserad
- Kamera-/bildbaserad ("Vision-based")
- Hybrid-baserad

Det finns åtskilliga **sensorbaserade tekniker** för att bestämma position och rörelse, bl a magnetiska (t ex kompass), akustiska, tröghetsensorer (accelerometer), optiska och mekaniska. Det finns även sensorer som bygger på olika typer av radiosignaler, t ex GPS och RFID.

Alla har sina fördelar respektive nackdelar med tanke på noggrannhet och felgenerering, t ex kan tröghetsensorer ge upphov till "drift" p g a brus-ackumulering.

**Kamerabaserade tekniker** använder sig enbart av videokameror för att bestämma position och rörelse. Första steget är att identifiera och "tracka" kännetecken ("features") i bilderna. Det kan vara allt ifrån enkla markörer som är utplacerade i förväg i rummet/omgivningen, till naturliga kännetecken

såsom hörn, kanter och kontraster som sedan bearbetas med lämplig metod för [bildanalys/bildbehandling](#) [42] för att den virtuella bilden ska kunna placeras rätt i den verkliga bilden.

Kamerabaserade system har lågt ”jitter” och ingen ”drift” och har fördelen att de kan korrigera fel dynamiskt.

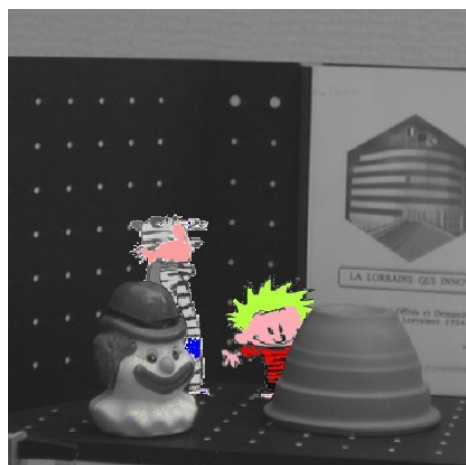
Kamerabaserade tekniker är dock långsamma, och snabba kamerarörelser kan leda till att trackingen misslyckas, särskilt i de system som använder punkter och strukturer som stöd för trackingen, och det kan ta tid för systemet att rätta till detta. Därför kombineras ofta kamerabaserade system med t ex GPS och accelerometrar till **hybrid-system**, som kompenserar för långsamheten i kamerabaserade tekniker.

De senaste åren har det kommit Realtids tracking-system som kan hantera rörelseoskärpa (”motion-blur”). Det har visat sig att rörelseoskärpa i kamerabaserade system kan analyseras och användas till att ersätta t ex ett gyroskop [9].

## Problem vid tracking

Tracking-fel i AR-system kan också orsakas av t ex en komplex miljö, ljusändringar, rörelser i miljön, eller att objekt i rörelse separeras eller slås samman p g a ”occlusion”.

Ett problem som kan uppstå i AR-system är felaktig ”occlusion”. Virtuella objekt ska ibland gömmas bakom reella objekt, och reella objekt kan döljas av virtuella. Virtuella objekt placeras dock alltid ”överst”, dvs framför reella objekt, även om det reella objektet ska vara närmare kameran. Detta är inte trovärdigt, och mycket viktigt att det är korrekt i seriösa applikationer, t ex inom reparation och medicinska applikationer. Det är stor risk att problem uppstår vid användande av en (1) kamera, då man inte använder en 3D-modell av den reella miljön eller objektet. En lösning är att använda två kameror som genererar ”depth-maps”. Det finns också lösningar då en (1) kamera används, utan en 3D-modell; t ex kan ”urklippsmasker” genereras baserat på rörelsen på konturer i bilden över tiden [9].



Figur 9. ”Occlusion” i AR.

## Kamerabaserad tracking

För att kunna placera ett 3D-objekt korrekt i den kombinerade virtuella och reella verkligheten som återges på displayen i en AR-applikation, krävs det att kamerans position och rörelser kan detekteras. För att upplevelsen ska bli trovärdig, måste tracking-tekniken vara robust och tåla t ex snabba rörelser utan att ”jitter” och ”lag” uppstår mellan det virtuella objektet och den reella verkligheten. Detta är ju inte minst viktigt inom medicinska applikationer eller vid reparation av viktiga objekt med hjälp av AR. Jämför man med VR, är det mycket lättare för användaren att upptäcka fel i AR, eftersom man har den reella verkligheten som referens.

Det första steget i tracking-proceduren är att identifiera kännetecken (”features”). En feature är en speciell punkt i bilden som tracking-algoritmen kan ”läsa på” och följa genom flera bildrutor. Som feature väljs ofta ljusa eller mörka punkter, kanter eller hörn, beroende på vilken tracking-algoritm som används. Det viktiga är att varje feature representerar en specifik punkt på det verkliga objektets yta. När en feature är identifierad och ”trackad” skapas en serie med tvådimensionella koordinater, som representerar ”features” position över en serie av bildrutor. Det är denna serie som bildar ett ”track”. När väl dessa ”tracks” har skapats kan de omedelbart användas för 2D-”[motion tracking](#)”, och användas till att beräkna 3D-information [6, 46].

För att göra det möjligt att återge de virtuella objekten exakt rätt i bilden av den reella verkligheten, krävs det alltså att det finns eller skapas en motsvarande virtuell representation av verkligheten, där de virtuella objekten placeras, samt att denna modell/karta följer kamerans rörelser och position.

Det finns flera olika kamerabaserade tracking-metoder [3]:

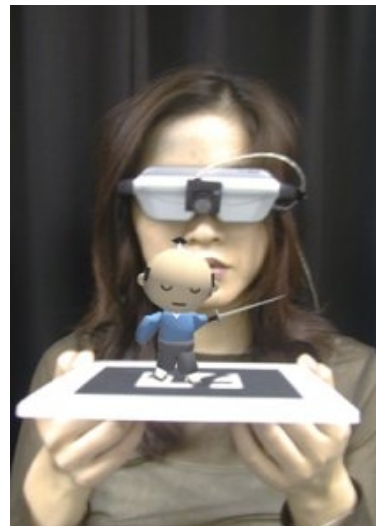
- Markör-baserad
- Modell-baserad
- ”Natural features”

## Markörer

Markörer kan vara passiva eller aktiva. Exempel på aktiva markörer är LEDs. Passiva markörer kan bestå av cirklar (t ex utskrivna på papper) som är placerade på kända ställen i miljön och fungerar som referenspunkter, eller kvadratiske kort med olika unika symboler eller mönster.

Till skillnad från cirkelmarkörerna krävs det endast ett kvadratisk markörkort för att kameran ska kunna registrera positionen på markörens fyra hörn i förhållande till varandra, och det virtuella objektet kan därmed återges i rätt vinkel och storlek. Vilket objekt som återges är kopplat till det unika mönstret på kortet. Mönstret eller symbolen på kortet identifieras genom linje- och kontrastdetektering.

Ett exempel på ett system med kvadratiske markörer är [ARToolKit](#) [24], som presenterades redan år 1999 [27], men fortfarande är populärt i många AR-applikationer. Andra sådana system är IGD (Institut Graphische Datenverarbeitung), SCR (Siemens Corporate Research) och HOM (Hoffman marker system) [14] (se figur 11).



Figur 10. [ARToolKit](#) [24].



Figur 11. Markörer från olika system: [ARToolKit](#), [IGD](#), [HOM](#) respektive [SCR](#) [14].

En nackdel med markör-baserade tekniker är att de inte är skalbara (t ex utomhus).

I figur 12–14 och i motsvarande videor visas exempel på tillämpningar av markörbaserad AR.



Figur 12: [ARf](#), ett Augmented Reality-virtuellt husdjur i iPhone, framtagen av [Augmented Environments Lab](#); en forskningsgrupp på [GVU Center](#) vid [Georgia Institute of Technology](#) [18].



*Figur 13. En video som [demonstrerar ARToolkit](#) [17].*



*Figur 14. En annan demonstrationsvideo av markör-baserad AR: "[Mini-tanks came out in sequence from a hole in my room wall](#)" [36].*

### Markörer med "natural shapes"

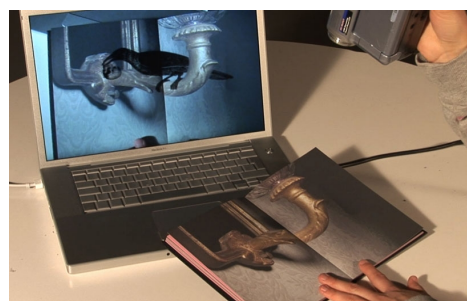
Det finns framtaget tekniker där man istället för att ha olika mönster eller grafiska symboler på markörkortet, använder figurer i form av naturliga former ("natural shapes"). [Visual Media Lab](#) på [Ben Gurion University](#) [40] har i samarbete med [Mark Billinghurst](#) på [HIT Lab NZ](#) [25] framställt några exempel med denna nya teknik [21]. Man använder här en typ av "computer-vision"-teknik för "feature-tracking" som benämns "[Blob-detection](#)" [43, 45] (se figur 15 och tillhörande video). Tekniken finns redan framtagen för iPhone [21].



*Figur 15. En demonstration av [markörer med "natural shapes"](#) [21].*

### Utan markörer

"[The Haunted Book](#)" [12] är en bok där man kompletterar bilderna i den fysiska boken med animeringar genom AR-teknik. Man har dock inte använt sig av markörer för framställningen av den förstärkta verkligheten, som t ex Mark Billinghurst m fl gjort i tidigare liknande arbeten med "Magic Books" [4], utan istället använt de "naturliga" bilderna på sidorna i boken (se figur 16 och demonstrationsvideo).



*Figur 16. Videon som visar "[The Haunted Book](#)" [12].*

## Modell-baserad tracking

I denna metod utgår man från en CAD-modell (eller en 2D "template") av den reella miljön eller objektet [3]. Olika bildanalysmetoder används sedan för att identifiera "features" i bilden; hörn, kanter, linjer, former och strukturer, som jämförs med CAD-modellen för att den virtuella bilden fortlöpande ska kunna positioneras korrekt i den reella bilden, efter kamerans rörelser.

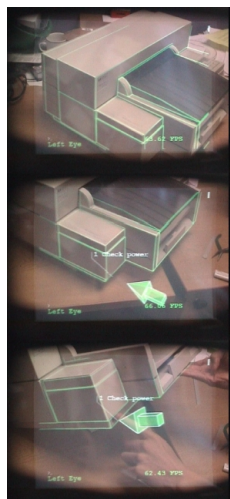
Två huvudmetoder finns; kant-baserade ("edge-based") samt struktur-baserade tekniker ("textured-based") [11]. Dessa metoder går att kombinera.

Kantmetoden är robust mot ljusändringar och är drift-fri. Fel kan dock uppstå genom att "trackern" låser mot fel kant, vilket resulterar i "tracking failure". Genom att kombinera med en metod för "point features" kan man få en mer robust metod [9].

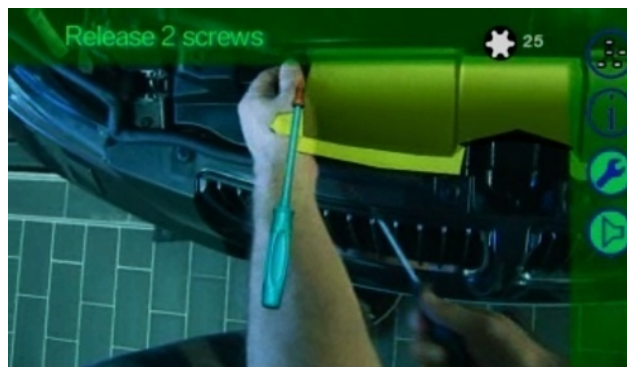
Vid struktur-metoden kan det uppstå problem vid skal- och ljusförändringar [11].

Nackdelen för samtliga modell-baserade metoder är att det måste konstrueras en modell i förväg, innan tracking kan ske [3].

Exempel på användningsområde för modell-baserade metoder är t ex när man ska utföra en uppgift, exempelvis en reparation [19] (se figur 17 och 18 samt till figur 18 tillhörande video). I figur 19 och 20 och i motsvarande videor visas exempel på tillämpningar av modell-baserad AR.



**Figur 17.** Reparations- och underhålls-instruktioner med hjälp av AR [9].



**Figur 18.** Reparationsinstruktioner för en BMW-motor med hjälp av AR. Ett forskningsprojekt av BMW [19].



**Figur 19.** Demo I av modell-baserad tracking, framtagen av [Lagadic research group](#) vid [Irisa/INRIA](#) [33].



**Figur 20.** Demo II av modell-baserad tracking, framtagen av [Lagadic research group](#) vid [Irisa/INRIA](#) [33].



## ”Natural features” och SLAM

Även i metoder som enbart nyttjar ”natural features” används en modell av den reella verkligheten för att kunna positionera den virtuella bilden, men skillnaden är att det finns inte en färdig modell från början, utan denna skapas efter hand genom att ”tracka features” i bilden; punkter eller delar av objekt.

Metoden bygger på två steg [3]:

1. Registrering av features (”learning stage” – låg beräkningsintensitet).
2. ”Tracka” dessa features i videoflödet (”training stage” – hög beräkningsintensitet).

Metoder som bygger på identifikation av ”natural features” utan tidigare vetskap om hur omgivningen ser ut, har sitt ursprung i robotvärlden där metoderna används för att bygga upp kartor av den okända omgivningen, t ex använde robotfordonen Mars Exploration Rovers som landsattes på planeten Mars år 2004 en sådan metod [9, 43, 49] – SLAM (se nedan).

Frågan är till vilken nytta denna metod är inom Augmented Reality. Vad är det som ska förstärkas när man inte vet vad som finns i omgivningen? Vilken information kan man förvänta sig få när man i förväg inte vet vilka objekt som finns i miljön man rör sig i? Men sådana system kan fungera bra som ett komplement i miljöer som man bara delvis känner till, för att fylla i de ”tomrum” man eventuellt har [9]. En annan idé kan vara att anlita en expert via distans (”remote expert”) som kan förse miljön med virtuell förstärkning [7].

I metoden [SLAM](#) (Simultaneous Localization and Mapping) sker konstruktion av modeller från bildsekvenser, utan tidigare vetskap om hur miljön ser ut, eller endast vetskap av en liten del av miljön [49]. Man kan beskriva metodens två steg enligt följande:

- Mapping – hur ser världen ut?
- Localization – var är jag?

För att minimera felgenerering och brus (störningar) har man två separata men simultana processer för lokalisering och mapping. Det utförs iterativ återkoppling mellan processerna. För att ytterligare undvika fel som genereras av kamerarörelse/-hastighet, ”feature points” och den virtuella kamerans position används bl a Kalman filter [9, 49].

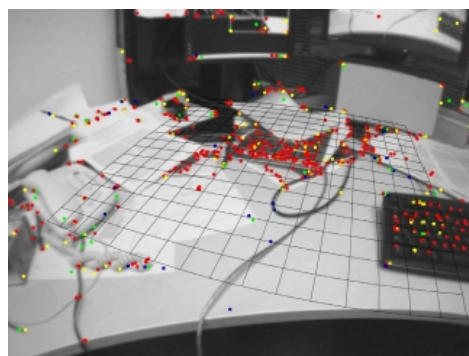
SLAM i sin ursprungliga form är mycket data- och beräkningsintensivt och kan inte användas på dagens smartphones [13]. Ett sätt att ”komma runt” detta är att överföra bildflödet till en arbetsstation via ett trådlöst nätverk och betrakta PDA:n som en klient [9, 13], men man mister då delvis tanken om mobil AR.

## PTAM

[Georg Klein](#) [29] och David Murray på University of Oxford har utvecklat ett alternativ till SLAM, som benämns PTAM – Parallel Tracking and Mapping. Metoden, med vilken man kan utföra samtidig ”tracking” och ”mapping” i en tidigare okänd miljö, delar upp dessa två uppgifter att utföras parallellt på en dator försedd med en dubbelkärnig processor. Med PTAM kan tusentals landmärken ”trackas” i realtid med en noggrannhet och robusthet jämförbar med modell-baserade system [7].

När en karta väl är skapad, kan ett virtuellt horisontalt plan etableras, på vilket det är möjligt att placera in virtuella objekt (se figur 21).

För att användaren ska kunna interagera i denna miljö i realtid med en handhållen kamera, krävs det att



**Figur 21.** Ett virtuellt horisontalt plan är skapat, på vilket det är möjligt att placera in virtuella objekt.

trackingen är snabb, noggrann och robust, samtidigt som kartan förfinas och växer när nya områden tillkommer. För att åstadkomma detta har Klein valt att begränsa den okända miljön i storlek till ”small AR Workspaces” samt att miljön i huvudsak är statiskt. Med ”liten miljö” menas i detta sammanhang ett skrivbord, ett normalstort rum, eller utomhus på en begränsad yta, t ex framför en byggnad. I dagsläget stödjer PTAM-metoden inte att man ”springer runt på stan”.

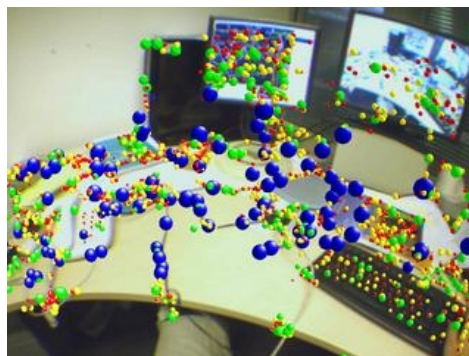
Som nämnts ovan separeras tracking och mapping, och uppgifterna utförs i två separata ”trådar”. Mappingen baseras på ”keyframes”, och utförs genom batch-teknik. Nya trackingpunkter läggs till genom [epipolar-sökning](#) [44], och tusentals punkter i olika storleksnivåer registreras (se figur 22).

Skillnaden mot monokulära SLAM-metoder för mobil AR, som ju också utför en samtidig process med tracking och mapping, är att i SLAM-metoderna uppdateras registreringen av kamerans position och varje landmärke samtidigt i varje enskild bildruta. När en snabbbrörlig handhållen kamera används med sådana SLAM-metoder, är risken stor att det uppstår problem i mappingen, vilket gör att de inte anses vara tillräckligt robusta i sammanhang med mobil AR [7].

Genom PTAM-metoden är trackingen inte lika hårt knuten till mapping-proceduren, och därmed mer användbar för mobil AR, eftersom kartan inte måste uppdateras i varje bildruta. Detta är möjligt eftersom kartan oftast ändå inte behöver ändras i varje bildruta, p g a att tidvis är kameran mer eller mindre stilla, vilket innebär att keyframes kan utnyttjas för mappingen, och bildanalysen kan därmed utföras mer genomarbetad.

I [demonstrationsfilmen](#) som Klein satt samman [29, 30] visas exempel på flera av PTAM-metodens styrkor:

- En avsiktlig skakning av kameran gör att trackingen tillfälligt förloras, men systemet återhämtar sig snabbt.
- Mappingen pågår i bakgrunden, utan att användaren märker detta vid interaktionen.
- Den handhållna kameran fungerar både som display och kontroll-/inmatningsenhet i både DarthVader-spelet och brännglas-sekvensen (se figur 23), och visar systemets precision samt tålighet mot snabba rörelser.
- Skalförändringar fungerar utan problem för tracking-systemet – man kan röra kameran väldigt nära ett objekt och sedan förflytta sig långt ifrån (inom det begränsade utrymmet) med bibehållen precision i trackingen. Rörelsen kan göras mycket snabbt, utan att systemet förlorar trackingen p g a rörelseoskärpa (”motion blur”). I SLAM-baserade system uppstår lätt problem vid stora skalförändringar och snabba rörelser.



**Figur 22.** En mängd registrerade ”features” i form av punkter.



**Figur 23.** Bilder från Kleins [demonstrationsfilm](#) av PTAM [29, 30].

Dock blir det i PTAM-metoden problem för mappingen när man skalar upp miljön, dvs när det ursprungliga ”small workspace” blir för stort. När kartan blir för stor får systemet problem med att lägga till nya keyframes och punkter till kartan. Det blir för beräkningsintensivt att hålla reda på trackingpunkterna i en alltför stor karta i ett system för mobil AR. Dock fungerar trackingen fortfarande i realtid. Klein har lyckats skapa en karta över ett kontor i 360° och som består av 11.000 kartpunkter och 280 keyframes. En mer praktisk gräns för systemet är ca 6.000 punkter och 150 keyframes [7].

En anledning till att använda många features, som i PTAM, är att det bidrar till att reducera jitter och dessutom förbättrar systemets prestanda när några features är dolda eller felaktiga.

I SLAM-baserade system minskar bildfrekvensen (”frame-rate”) när miljön blir för stor, till skillnad från PTAM, där bildfrekvensen alltså inte påverkas, men hastigheten med vilken nya delar av omgivningen kan utforskas minskar.

PTAM-metoden är inte felfri. Det förekommer tracking-fel och felmapping. Metoden klarar av tillfällig ”occlusion”, men inte när det sker stora förändringar av den kända miljön, dvs när miljön inte längre i huvudsak är statisk. En annan nackdel med PTAM är att kartan endast består av ett ”punkt-moln”. Efter att det horisontala planet etablerats, där de virtuella objekten placeras in, fungerar kartan enbart som ett verktyg för kamera-trackingen. Systemet drar inga slutsatser om omgivningens geometri. Detta innebär bl a att reella objekt som ligger utanför det etablerade horisontala planet inte påverkar de virtuella objekten, t ex genom att de virtuella figurerna stoppas när de ”springer på” dessa reella objekt eller döljs bakom dem. Detta är ett ämnesområde att forska vidare inom.

Det finns en [kompletterande film](#) med ytterligare demonstration av PTAM i funktion i utomhusmiljöer samt i en större kontorsmiljö [29, 31].

## PTAM på iPhone

Processorerna i smartphones och mobiltelefoner med kamera har de senaste åren utvecklats så att de nu klarar beräkningsintensiva uppgifter som kamerabaserad tracking. Men det återstår en del utveckling av den övriga ingående hårdvaran som krävs för att kamerabaserad tracking ska fungera utan problem; bildvinkeln på mobilkamerornas objektiv är alltför snäv för att trackingen ska bli robust, bildfrekvensen är låg och slutartiden lång vilket lätt leder till rörelseoskärpa, och en ”rolling shutter” kan orsaka skeva och snedvridna motiv [48].

Klein har anpassat PTAM-metoden så att mjukvaran kan användas på en kameramobil, och då mer specifikt en iPhone 3G [8]. Eftersom iPhones enklarkärniga processor är 15–30 gånger långsammare än en (1) processorkärna i en ”vanlig” dubbelkärnig PC-processor, blir den möjliga storleken på kartan som skapas från den okända miljön, mycket mindre än då PTAM körs på en PC. Dessutom är det av samma anledning inte möjligt med PTAM i en kameramobil att tracka en så stor mängd ”point fetatures” i varje bildruta, och därmed inte heller på flera storleksnivåer. Tracking-metoden för PTAM-systemet för mobiltelefoner kräver också att miljön som ”trackas” innehåller mer tydliga mönster och strukturer än då det används i en PC.

Tillförlitligheten och noggrannheten når i dagsläget naturligtvis inte upp till samma nivå som när PTAM körs på PC tillsammans med en handhållen kamera, men [demonstrationsvideon](#) [29, 32] (se figur 24) visar att det är möjligt att använda t ex en iPhone för



*Figur 24. Kleins demonstrationsfilm av [PTAM på en iPhone](#) [29, 32].*

kamerabaserad tracking i en okänd miljö, med syftet att kunna placera in virtuella objekt i den reella miljön på ett trovärdigt sätt, och därmed skapa små AR-miljöer, både inom- och utomhus.

Jag är övertygad om att vi får se en mängd nya ”äkta” (dvs som följer Azumas definition) AR-applikationer för smartphones och kameramobiler inom en snar framtid, när hårdvaran för kamerabaserad tracking ytterligare utvecklats, och också i kombination med kameramobilernas övriga sensorer som GPS, accelerometer och kompass.

Lars Peterson, 2010-01-22.

[lars@flexibel.se](mailto:lars@flexibel.se)

## Referenser

- [1] Andersson, A., 2009. *GPS-mobiler ger plats för helt nya spel*. Metro teknik, 25 november – 1 december 2009, sidorna 12–13.
- [2] Azuma, R., Bailiot, Y., Behringer, R., Feiner, S., Julier, S., MacIntyre, B., 2001. *Recent advances in augmented reality*. Computer Graphics and Applications, IEEE. Publication Date: Nov.-Dec. 2001. Volume: 21, Issue: 6. On page(s): 34 – 47.
- [3] Feng, Z., Duh, H.B.-L., Billinghurst, M., 2008. *Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR*. 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, 2008. ISMAR 2008. On page(s): 193 - 202.
- [4] Grasset, R., Dunser, A., Billinghurst, M., 2008. *The design of a mixed-reality book: Is it still a real book?* 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, 2008. ISMAR 2008. On page(s): 99 – 102.
- [5] Holm, N., 2009. *En ny värld på väg in i mobilen*. Sydsvenskan, 2009-12-18, sidorna A22–A23. *En ny värld på väg in i mobilen*: <https://www.sydsvenskan.se/2009-12-18/en-ny-varld-pa-vag-in-i-mobilen>  
*Få tjänster färdiga i Sverige*: <https://www.sydsvenskan.se/2009-12-18/fa-tjanster-fardiga-i-sverige>  
*Forskaren tror tekniken kommer att blomstra*: <https://www.sydsvenskan.se/2009-12-18/forskaren-tror-tekniken-kommer-att-blomstra>
- [6] Kato, H., Billinghurst, M., 1999. *Marker tracking and HMD calibration for a video-based augmented reality conferencing system*. 2nd IEEE and ACM International Workshop on Augmented Reality, 1999. (IWAR '99). On page(s): 85 - 94.
- [7] Klein, G., Murray, D., 2007. *Parallel Tracking and Mapping for Small AR Workspaces*. 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, 2007. ISMAR 2007. On page(s): 225 - 234.
- [8] Klein, G., Murray, D., 2009. *Parallel Tracking and Mapping on a Camera Phone*. 8th IEEE and ACM International Symposium on Mixed and Augmented Reality, 2009. ISMAR 2009.
- [9] Klein, G., 2006. *Visual Tracking for Augmented Reality*. PhD Thesis, University of Cambridge.
- [10] Mjömark, P-O., 2009. *Kunskap försvinner med medarbetarna*. Metro teknik, 4–10 november 2009, sidan 15.
- [11] Pressigout, M., Marchand, E., 2006. *Hybrid tracking algorithms for planar and non-planar structures subject to illumination changes*. IEEE/ACM International Symposium on Mixed and Augmented Reality, 2006. ISMAR 2006. On page(s): 52 - 55.
- [12] Scherrer, C., Pilet, J., Fua, P., Lepetit, V., 2008. *The haunted book*. 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, 2008. ISMAR 2008. On page(s): 163 - 164. *The haunted book*: [http://chipchip.ch/CVMID3\\_1/haunted\\_book.html](http://chipchip.ch/CVMID3_1/haunted_book.html)

- [13] Wagner, D., Schmalstieg, D., 2009. *History and Future of Tracking for Mobile Phone Augmented Reality*. International Symposium on Ubiquitous Virtual Reality, 2009. ISUVR 2009. On page(s): 7 - 10.
- [14] Xiang Z., Fronz, S., Navab, N., 2002. *Visual marker detection and decoding in AR systems: a comparative study*. International Symposium on Mixed and Augmented Reality, 2002. ISMAR 2002. On page(s): 97 - 106.
- [15] Wallergård, M., Eriksson, J., 2008. *Virtual Reality in Theory and Practice*. Kurskompendium, Lunds Universitet. Kapitel 10; Interfaces for augmented reality applications, sidorna 70 - 72.
- [16] ACM Digital Library (Association for Computing Machinery). <http://portal.acm.org/portal.cfm>
- [17] *ARToolkit demonstration*:  
<http://www.youtube.com/watch?v=TqGAqAFIGg0>
- [18] Augmented Environments Lab, a research group in the GVV Center at the Georgia Institute of Technology.  
*ARf, an Augmented Reality Virtual Pet on the iPhone*.  
<http://www.augmentedenvironments.org/lab/2008/11/28/arf-iphone-peek/>  
[http://www.youtube.com/watch?v=\\_0bitKDKdg0](http://www.youtube.com/watch?v=_0bitKDKdg0)
- [19] BMW.  
*BMW Augmented Reality in practice. BMW research projects - workshop applications at BMW*.  
[http://www.bmw.com/com/en/owners/service/augmented\\_reality\\_workshop\\_1.html](http://www.bmw.com/com/en/owners/service/augmented_reality_workshop_1.html)  
<http://www.youtube.com/watch?v=P9KPJIA5yds>
- [20] Games Alfresco.  
*10 best augmented reality DEVICES that will reinvent mobile video games*.  
<http://gamesalfresco.com/2008/04/16/10-best-augmented-reality-devices-that-will-reinvent-mobile-video-games/>
- [21] Games Alfresco.  
*ISMAR 2009: Sketch and Shape Recognition Preview From Ben Gurion University*.  
<http://gamesalfresco.com/2009/10/16/ismar-2009-sketch-and-shape-recognition-preview-from-ben-gurion-university/>  
*Shape Recognition and Pose Estimation for Mobile Augmented Reality*:  
<http://www.youtube.com/watch?v=je8SihypYA4>
- [22] Ghostwire.  
<http://www.ghostwiregame.com>  
<http://www.adifferentgame.com>
- [23] GoWeb3D. <http://goweb3d.com>  
*GoWeb3D Augmented Reality on Layar Reality Browser*: <http://www.youtube.com/watch?v=aMoesX4POfk>

- [24] HITLab. The Human Interface Technology Lab.  
*ARToolkit*.  
<http://www.hitl.washington.edu/artoolkit/>
- [25] HIT Lab NZ. The Human Interface Technology Laboratory New Zealand.  
<http://www.hitlabnz.org>  
<http://www.hitlab.co.nz>
- [26] IEEE Xplore Digital Library. <http://ieeexplore.ieee.org>
- [27] Institute for Computer Graphics and Vision. Graz University of Technology.  
*History of Mobile Augmented Reality*.  
<https://www.icg.tugraz.at/~daniel/HistoryOfMobileAR/>
- [28] ISMAR. International Symposium on Mixed and Augmented Reality.  
<http://www.augmented-reality.org/ismar/>
- [29] Klein, G., 2009. Georg Klein Home Page. Active Vision Group. University of Oxford.  
<http://www.robots.ox.ac.uk/~gk/>
- [30] Klein, G., augsuti 2007.  
*Parallel Tracking and Mapping for Small AR Workspaces (PTAM)*:  
<http://www.youtube.com/watch?v=F3s3M0mokNc>
- [31] Klein, G., november 2007.  
*Parallel Tracking and Mapping for Small AR Workspaces (PTAM) – extra*:  
<http://www.youtube.com/watch?v=Y9HMn6bd-v8>
- [32] Klein, G., juni 2009.  
*PTAM + AR on an iPhone 3G*:  
<http://www.youtube.com/watch?v=pBI5HwitBX4>
- [33] Lagadic research group at Irisa/INRIA.  
*Real-time 3D localisation and tracking using virtual visual servoing. Markerless tracking. Integration within an animation module*.  
<http://www.irisa.fr/lagadic/demo/demo-ar3/demo-ar3.html>  
*Physics and Augmented Reality - Part 1*: <http://www.youtube.com/watch?v=enXTKvhE7yk>  
*Physics and Augmented Reality - Part 2*: <http://www.youtube.com/watch?v=umbTreYhidM>
- [34] Layar. <http://layar.com>  
*Layar, worlds first mobile Augmented Reality browser*:  
[http://www.youtube.com/watch?v=b64\\_16K2e08](http://www.youtube.com/watch?v=b64_16K2e08)
- [35] Layar.  
*Layar3D Augmented Reality Engine Trial; Venue Tagging*:  
<http://www.youtube.com/watch?v=rA4uB-Jg2Xo>

- [36] *Mini-tanks came out in sequence from a hole in my room wall:*  
<http://www.youtube.com/watch?v=oiqIPXnKkKo&NR=1>
- [37] Schmied, M., 2009. *24 hour business camp*. SweetAR, 2009-11-01.  
<http://sweetar.com/blog/?p=296>  
[http://2.bp.blogspot.com/\\_nmz1KfNpgos/Suq5tpP\\_4GI/AAAAAAAAAAk/SmSeKxolnb0/s1600-h/Speech\\_START.jpg](http://2.bp.blogspot.com/_nmz1KfNpgos/Suq5tpP_4GI/AAAAAAAAAAk/SmSeKxolnb0/s1600-h/Speech_START.jpg)
- [38] TAT – The Astonishing Tribe. <http://www.tat.se>  
*TAT UI Prototypes:* [http://www.tat.se/site/showroom/latest\\_design.html](http://www.tat.se/site/showroom/latest_design.html)  
*TAT augmented ID:* <http://www.youtube.com/watch?v=tb0pMeg1UN0>
- [39] UgoTrade. <http://www.ugotrade.com/>
- [40] Visual Media Lab at Ben Gurion University.  
<http://www.vml.cs.bgu.ac.il>  
<http://www.cs.bgu.ac.il>
- [41] Wikipedia. *Augmented Reality*.  
[http://en.wikipedia.org/wiki/Augmented\\_reality](http://en.wikipedia.org/wiki/Augmented_reality)
- [42] Wikipedia. *Bildanalys (bildbehandling)*.  
[http://sv.wikipedia.org/wiki/Bildanalys\\_\(bildbehandling\)](http://sv.wikipedia.org/wiki/Bildanalys_(bildbehandling))
- [43] Wikipedia. *Computer vision*.  
[http://en.wikipedia.org/wiki/Computer\\_vision](http://en.wikipedia.org/wiki/Computer_vision)
- [44] Wikipedia. *Epipolar geometry*.  
<http://en.wikipedia.org/wiki/Epipolar>
- [45] Wikipedia. *Feature detection*.  
[http://en.wikipedia.org/wiki/Feature\\_detection\\_\(computer\\_vision\)](http://en.wikipedia.org/wiki/Feature_detection_(computer_vision))
- [46] Wikipedia. *Match moving*.  
[http://en.wikipedia.org/wiki/Match\\_moving](http://en.wikipedia.org/wiki/Match_moving)
- [47] Wikipedia. *Picture-in-picture*.  
<http://en.wikipedia.org/wiki/Picture-in-picture>
- [48] Wikipedia. *Rolling shutter*.  
[http://en.wikipedia.org/wiki/Rolling\\_shutter](http://en.wikipedia.org/wiki/Rolling_shutter)
- [49] Wikipedia. *Simultaneous localization and mapping*.  
[http://en.wikipedia.org/wiki/Simultaneous\\_localization\\_and\\_mapping](http://en.wikipedia.org/wiki/Simultaneous_localization_and_mapping)
- [50] Wikipedia. *Virtuality Continuum*.  
[http://en.wikipedia.org/wiki/Virtuality\\_Continuum](http://en.wikipedia.org/wiki/Virtuality_Continuum)