

# **Kontinuerlig positionsbestämning med multipla kameror**

Linda Karlsson

Examensarbete  
Elektroteknik  
2010

EXAMENSARBETE	
Arcada	
Utbildningsprogram:	Elektroteknik
Identifikationsnummer:	2863
Författare:	Linda Karlsson
Arbetets namn:	Kontinuerlig positionsbestämning i realtid med multipla kameror
Handledare (Arcada):	Rene Herrman
Uppdragsgivare:	Arcada- Nylands Svenska Yrkeshögskola
<p>Sammandrag:</p> <p>Arbetets syfte var att forska i möjligheten att utveckla ett positionsbestämningssystem för tävlande hästar på en travbana genom att använda de kameror som redan är i användning där för bedömning av loppen. Arbetet utgår från de förhållanden som gäller på Vermo travbana i Helsingfors. Frågeställningen var om det är möjligt att bestämma positionen i världskoordinater för ett föremål med hastigheten 16 m/s tillräckligt snabbt och noggrant för att uppfylla krav på kontinuerlighet och realtid under ett travlopp. På grund av de höga hastigheterna och korta avstånden mellan kamerorna och objekten avgränsades arbetet till teknikerna datorseende och triangulering då dessa tekniker hade de teoretiskt bästa möjligheterna att genomföra uppgiften. För att bestämma snabbheten på processen som skall utföras delades den upp i tre delmoment och varaktigheten för varje moment bestämdes. Det första delmomentet var kameran tar en bild begränsades av videoformatets uppdateringshastighet som på Vermo travbana var PAL formatet med 25 bilder i sekunden. För det andra delmomentet med detektering och identifiering av objekten i bilden vilket representerar tekniken datorseende utfördes en jämförelse mellan de två ansiktsdetekteringsalgoritmerna Viola-Jones och en Haar-kaskad samt ett test för cirkeldetektion med en Hough transform. Resultatet av en korrekt detektering och identifiering av objektet är en bildkoordinat i x-led vilken används för att bestämma en riktning från kameran mot objektet i form av en vinkel <math>\alpha</math>. Denna uträkning utgör det sista delmomentet som kallas triangulering. Resultatet av hela processen är världskoordinaterna för objektet på målområdet med en viss fördröjning från den absoluta realtiden och med en viss noggrannhet. Detta arbete ger en storleksordning på fördröjningen från realtid och noggrannheten för ett system med dessa utgångslägen.</p>	
Nyckelord:	Objektdetektion, positionsbestämning, kamera, triangulering
Sidantal:	41
Språk:	svenska
Datum för godkännande:	

DEGREE THESIS	
Arcada	
Degree Programme:	Electrical Engineering
Identification number:	2863
Author:	Linda Karlsson
Title:	Continuous positioning in real-time with multiple cameras
Supervisor (Arcada):	Rene Herrman
Commissioned by:	Arcada-Nylands Svenska Yrkeshögskola
<p>Abstract:</p> <p>The aim with this study is to investigate the possibilities to develop a positioning system for racing horses on a trotting racetrack using the cameras that already are in use there. This project takes off on the situation that takes place on Vermo racetrack in Helsinki. The research questions was if it is possible to determine the world coordinates of an object moving at 16 m/s fast and precise enough to fulfill the criteria of continuity in real-time during a trotting race. Because of the high speed of the object and the short distances between the cameras and the object, the study is limited to the techniques triangulation and Computer Vision while they had the theoretically best prospects to carry out the tasks required. To determine the rapidity of the process needed, it was divided into three elements so that the durance of each element could be defined. The first element occurs when the camera takes a picture and is restricted by the video format and its frame rate which were PAL and its 25 frames/s at Vermo racetrack. The second element where the object is detected and identified represents the technique Computer Vision. To define a time for this element a comparisment between the two face detection algorithms Viola-Jones and a Haar-cascade was done and a separate test for circle detection with a Hough transform. The result of a correct detection and identification of the object is a picture coordinate in the x direction that is used to determine a direction from the camera to the object as an angle <math>\alpha</math> which forms the last element of the process called triangulation. The result of the whole process with its three elements is the world coordinates of the object on the target area with a delay from the absolute real-time and an accuracy. This study gives a timeline for the smallest possible delay from real-time and best possible accuracy for a system with these starting-points.</p>	
Keywords:	Objectdetection, positioning, cameras, triangulation
Number of pages:	41
Language:	swedish
Date of acceptance:	

# Innehåll:

TACKORD .....	6
ORDLISTA.....	6
1 INTRODUKTION.....	8
1.1 Inledning.....	8
1.2 Bakgrund .....	8
1.3 Syfte och mål.....	11
1.4 Avgränsning.....	14
1.5 Metoder.....	14
2 DETEKTION OCH IDENTIFIERING.....	15
2.1 Inledning.....	15
2.2 Testtillfället för bestämning av bildkoordinater .....	17
2.3 Testresultat för bestämning av bildkoordinater .....	19
2.4 Slutsats för detektion och identifiering.....	24
3 POSITIONSBESTÄMNING MED TRIANGULERING.....	25
3.1 Inledning.....	25
3.2 Teorin bakom triangulering .....	25
3.3 Testtillfälle för triangulering.....	31
3.4 Slutsats för positionering med triangulering .....	37
4 DISKUSSION .....	38
KÄLLFÖRTECKNING.....	40

## Figurförteckning:

Figur 1: Vermo travbana och dess kameror. ....	10
Figur 2: Blockschema över delmomenten i processen. ....	11
Figur 3: Triangulering av ett objekt med två kameror. ....	13
Figur 4: Bildkoordinaterna i x- och y-led. ....	16
Figur 5: Negativa och positiva x-koordinater i en bild. ....	17
Figur 6: Jämförelse mellan Viola-Jones algoritmen och Haar-kaskaden. ....	21
Figur 7: Cirkeldetektion med olika bakgrunder. ....	23
Figur 8: De fysikaliska förhållandena mellan kamerans- och zoomens linser i en kamera. ....	26
Figur 9: Fysikaliska förhållandena för vinkeln $\beta$ . ....	27
Figur 10: De fysikaliska förhållandena för uträkningarna av vinkel $\alpha$ . ....	28
Figur 11: Tangens funktionen och dess relevans vid vinkeln $\alpha$ storlek. ....	29
Figur 12: Triangulering med två kameror. ....	30
Figur13: Testsituation för triangulering. ....	33
Figur 14: Jämförelse mellan de tre ekvationerna för uträkning av vinkeln $\alpha$ . ....	35
Figur 15: Det relativa felet av den linjära och den allmänna ekvationen jämfört med den teoretiska vinkeln. ....	36
Figur 16: Det absoluta felet av den linjära och den allmänna ekvationen. ....	36

## Tabellförteckning:

Tabell 1: Jämförelse över ansiktsigenkänning mellan Viola-Jones algoritmen och Haar-kaskaden. ....	20
Tabell 2: Testresultaten av cirkeldetektionen med en Hough transform. ....	22
Tabell 3: Förhållandena vid testet för triangulering. ....	32
Tabell 4: Den linjära och den allmänna ekvationens relativa och absoluta fel. ....	36

## TACKORD

Ett stort tack åt de personer som hjälpt till i detta projekt; Rene Herrman för handledningen, John Morelius för modifierandet av programkoderna för testtillfällena och Timo Koljonen för bakgrundinformationen om kameran systemet på Vermo travbana.

## ORDLISTA

Baudrat	beskriver den tid det tar från det att en signal tas emot i en mottagare tills mottagaren kan ta emot följande. Enheten är synonym med pulser per sekund.
CCU	Camera Control Unit är en enhet som gör det möjligt att fjärrstyra vissa egenskaper hos kameran. Enheten är placerad i kontrollutrymmena.
GPS	Global Positioning System är ett navigationssystem bestående av 32 satelliter som kretsar runt jorden. Signaler från minst fyra satelliter krävs för att en position på jordytan skall vara möjlig att bestämmas.
Haar-kaskad	är ett degenererande beslutsträd av pixelsummor från delområden i en bild. Pixelsummorna påvisar om en sökt egenskap finns i det området eller inte.

Hough Transform      en matematisk process som hittar linjer och figurer i bilder genom att fastställa en pixels värde utgående från den själv och dess grannar. Om pixeln innehåller en kant används detta värde som parameter i den matematiska processen.

SDI                      Serial Digital Interface är en överföringsmetod för digitala ickekomprimerade och okrypterade videosignaler. Överföringshastigheten för detta gränssnitt varierar mellan 143 och 360 Mbit/s för videoformatet PAL.

# 1 INTRODUKTION

## 1.1 Inledning

Travsport är en internationell sport med utövare främst i Europa, Nordamerika, Australien och Nya Zeeland. I Finland finns det 43 travbanor som kör sammanlagt 560 trav om året vilka samlar in 200 miljoner euro genom olika totalisator spel (Suomen Hippos, 2009).

Tekniken som används för att bedöma travloppen i Finland utgörs endast av videokameror placerade på olika ställen runt banan. Detta leder till att dömandet av hästarnas rörelser på banan utgörs av domarnas iakttagelseförmåga vilket väcker mycket diskussion och hård kritik.

Detta arbete forskar i möjligheten att utveckla ett positionsbestämningssystem som kan bestämma hästarnas rörelser på banan under loppets gång med hjälp av de kameror som redan finns där. Teknikerna som forskas i detta arbete är datorseende och triangulering. Grundiden är att använda flera stationära kameror för att bestämma en precis position av hästarna i världskoordinater. Projektet utgår från de förhållanden som gäller på Vermo travbana i Helsingfors.

## 1.2 Bakgrund

Storleken på hästarnas positionsförändringar under travloppen ligger inom en ram från några centimeter till en meter vilket sätter kravet på noggrannhet för detta system till en centimeter för att ett positionsbestämningssystem skall vara av nytta.



Hästarna rör sig normalt under ett lopp med en hastighet på 40-50 km/h medan den momentana maximala hastigheten kan uppgå till 60 km/h. Detta betyder att det är möjligt att en häst rör sig med en hastighet på 16 m/s, vilket betyder att det tar endast 0,625 ms för den att förflytta sig en centimeter. Även om hastigheter på 16 m/s är mycket momentana och ovanliga är de dock möjliga.

De kameror som används idag på Vermo travbana är digitala videokameror av märket Panasonic AW-E860L (Panasonic, 2008) med zoomande motorobjektiv av märket Panasonic AW-LZ17MD9A 17X (Panasonic, 2006). Brännvidden på linserna varierar från 9 till 153 mm. Kamerorna styrs från kontrollutrymmena i domartornet med en styrspak. De zoomas in och ut enligt behov och kan svängas i sidled 180°.

Kamerorna och deras zoomande objektiv styrs genom en CCU-kontrollenhet av märket Panasonic AW-CB400L för maximalt fem kameror. Kontrollenheten används för fjärrkontrollering av kamerornas parametrar som t.ex. zoom och rotering.

Överföringen från kamerorna till skärmarna sker genom SDI gränssnittet av Panasonic AW-BP504 SDI-kortet. Denna överföringsmetod används för okomprimerade och okrypterade digitala videosignaler som överförs över sträckor kortare än 300 m (Wikipedia, 2010b).

Bildförhållandet som används under travloppen är 16:9. Mängden pixlar i en bild uppgår till 595728 stycken av vilka 538200 är effektiva element vilket horisontellt ger 960 pixlar och vertikalt 540 pixlar. Då kamerans bildyta är  $51,84 \text{ mm}^2$  ( $9,6 \text{ mm} * 5,4 \text{ mm}$ ) är storleken på en pixel  $87 \mu\text{m}$ . Brännvidden för kamerans lins är 15,24 mm.

Videoformatet som är i användning på Vermo travbana är PAL vilket har en uppdateringshastighet på 25 bilder i sekunden. Denna uppdateringshastighet leder till att det tar 40 ms för kameran att ta en bild. Om hästen då springer med en hastighet på 16 m/s hinner den röra sig 640 mm innan kameran tar följande bild.

Kamerorna är placerade på olika ställen runt banan. De två kameror som i dagens läge utgör tyngdpunkten i dömandet av travloppen är belägna i respektive framhorn av taket på domartornet. De två andra kamerorna som används är belägna vid ytterkanten av banan i vardera kurvan. Kamerornas placeringar runt banan beskrivs i figur 1.



*Figur 1: Vermo travbana och dess kameror.*

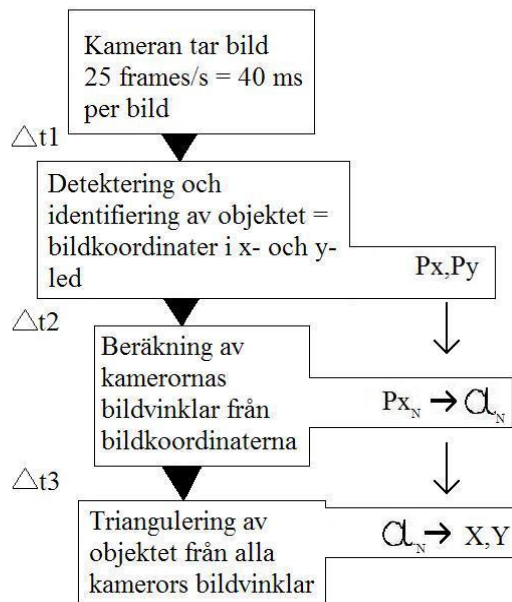
*1. Takkamera nr 1, 2. takkamera nr 2, 3. kameran i stallkurvan, 4. kameran i den bortre kurvan. Bilden är tagen från Google Earth och kamerornas placeringar är märkta av Timo Koljonen*

Banans längd är 1000 meter och alla travlopp i Finland körs motsols runt banan. Raksträckornas absoluta längder är 175 m och banans bredd är 24 m förutom startområdena på båda raksträckorna var banan är 26 m bred. Kurvornas radie är 85 m och deras lutning är 12,5°. Av hela banans längd utgör alltså kurvorna 650 m då vardera kurvan är 325 m lång. All information berörande Vermo travbana och dess kameror har Vermos representant Timo Koljonen framfört.

### 1.3 Syfte och mål

Kraven på detta system är kontinuerlighet under hela loppet, realtid för att göra bedömning under loppets gång möjlig, en noggrannhet på en centimeter, användarvänlighet och ekonomisk lönsamhet. Hästarna får inte belastas med stora föremål såsom sensorer eller andra prylar. Det får inte heller fästas andra föremål på hästarna som kan störa dem under loppet då minsta möjliga sak kan skrämma dem under en tävlingssituation. Sådana störningar kan orsakas av till exempel färggranna, fladdrande eller prasslande saker.

Den process som kommer att undersökas i detta projekt är beskriven i figur 2.



Figur 2: Blockschema över delmomenten i processen.

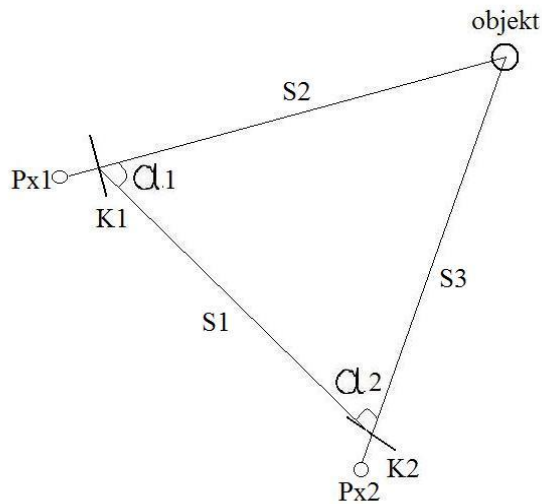
Tiden  $\Delta t_1$  är känd då den är beroende av kamerornas egenskaper medan  $\Delta t_2$  och  $\Delta t_3$  ännu är okända och deras storleksordning samt inverkan på den minsta möjliga fördröjning av realtiden skall bestämmas i detta projekt.

Tiden  $\Delta t_1$  representerar den tid det tar för kameran att ta en bild, det vill säga 40 ms. Denna tid går inte att påverka då den är beroende av videoformatets uppdateringshastighet. Tiden som kameran behöver för att ta en bild orsakar en tidslucka då det inte finns information om vad som händer på banan. Detta betyder att det inte finns någon information om hästens rörelser på banan under de 640 mm som den hinner röra sig under de 40 ms då kameran tar en bild.

Tiden  $\Delta t_2$  beskriver tiden det tar för en algoritm att hitta och identifiera hästen i bilden samt bestämma dess bildkoordinater i x- och y-led. Den bildkoordinat som är relevant för att kunna bestämma hästens position är den i x-led som berättar om hästens position på marken medan y-koordinaten inte är lika signifikant då hästens höjd inte påverkar dess position på marken.

Tiden  $\Delta t_3$  representerar tiden det tar att beräkna vinkeln  $\alpha$  från kameran mot hästen och därifrån bestämma hästens position i världskoordinater. För att räkna ut vinkeln  $\alpha$  behövs det x-koordinaten i bilden och den zoomande linsens brännvidd. Med vinkeln  $\alpha$  kan man sedan räkna ut sträckan från kameran till objektet.

Objektets position i världskoordinater kan bestämmas då sträckorna och vinklarna mellan alla kameror och objektet är kända. Denna situation är upplagd i figur 3.



Figur 3: Triangulering av ett objekt med två kameror.

Då delmomenten i figur 2 är utförda börjar processen om från början med följande bild från kamerorna. Resultatet av många korrekta positionsberäkningar skall vara en graf som presenterar de uträknade koordinaterna som om kamerorna filmat banan rakt uppifrån.

Den tid som processen i figur 2 förbrukar leder till en fördröjning från den absoluta realtiden. Vid utgångsläget är denna fördröjning 40 ms ( $\Delta t_1$ ) och den slutliga fördröjningen bestäms av alla tre deltider. För att bestämma tiden  $\Delta t_2$  utfördes det ett test med två ansiktsdetektionsalgoritmer och ett test med en cirkeldetektionsalgoritm varav alla tre kan utföra en snabb och säker detektion och identifiering av objekt i videosekvenser. Då det tredje delmomentet i figur 2 utgör förutom en tid  $\Delta t_3$  också det planerade systemets noggrannhet har det utförts ett test med triangulering som ger tiden  $\Delta t_3$  en längd och sätter krav för att en noggrannhet på en centimeter skall uppnås.

## 1.4 Avgränsning

Detta arbete koncentrerar sig på att utforska möjligheten till ett positionsbestämningssystem med datorseende och triangulering för att dessa tekniker har de bästa möjligheterna att uppfylla kraven som ställts på systemet. Andra tekniker som skulle ha varit mer logiska för ett positionsbestämningssystem förkastades då de inte kunde uppfylla snabbhetskraven vilka är essentiella.

En GPS eller en differentiell GPS applikation skulle vara lätt att implementera men på grund av deras långsamma baudrat på 1 s måste dessa tekniker avgränsas från detta projekt. En baudrat på 1s och ett mobilt objekt med en hastighet på 16 m/s skulle ge systemet en noggrannhet på 16 m.

## 1.5 Metoder

För att få den behövliga teorin bakom triangulering och datorseende studerade jag olika källor. Information om triangulering fanns i tillräckliga mängder i bokform även om det visade sig att uträkningarna kan utföras med vanlig matematik som man hittar i vilka matematikböcker som helst. Datorseende är en nyare teknik om vilken man hittar bra information om på Internet i form av artiklar.

Den fria encyklopedin Wikipedia på Internet använde jag endast för att snabbt förklara vissa tekniska termer och matematiska begrepp för att underlätta förståendet av texten.

Tidigare arbeten med liknande kombinationer av tekniker var svåra att hitta då de flesta system av denna natur används av industriella företag som inte publicerar sina implementationer.

För att komma fram till konkretare tidsaspekter för tiderna  $\Delta t_2$  och  $\Delta t_3$  utförde jag vissa tester och uträkningar. Med resultaten från testet för trianguleringen kunde man dessutom bestämma vissa krav för att en noggrannhet på en centimeter i positionsbestämningen skall uppstå.

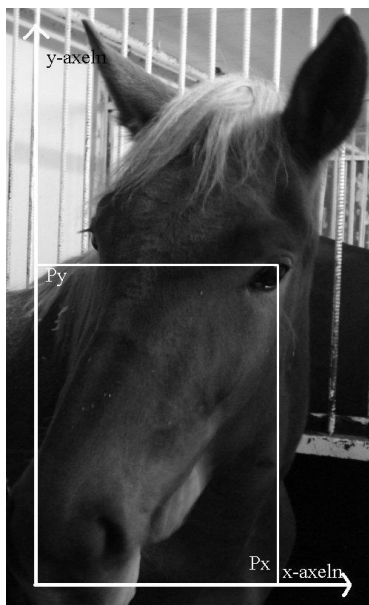
## 2 DETEKTION OCH IDENTIFIERING

### 2.1 Inledning

Med en detektion av en häst menar man att programmet hittar en häst i bilden medan identifieringen bestämmer vilken häst som syns i bilden. Hästarnas antal i ett travlopp kan uppgå till 16 stycken och de skiljs åt genom nummerlappar som är fästa på vardera sidan av hästen. Kraven på både detektionen och identifieringen är att processerna skall ske så snabbt som möjligt för att inte nämnvärt försämra den minsta möjliga fördröjningen från realtid på 40 ms som bestäms av kamerorna.

I det planerade systemet skall programmet kunna skilja åt alla de 16 möjliga ekipage av hästar och kuskar från varandra genom en identifierande egenskap på hästen. För en identifiering med datorseende från flera olika ställen på en gång krävs det en identifierande egenskap som till sitt utseende ser lika ut från alla håll. Därför är det befogat att anta att nummerlapparna som idag är i användning inte kunde användas till identifiering av hästar i det system som planeras här.

Resultatet av en korrekt detektion och identifiering skall vara bildkoordinaterna i x- och y-led i bilderna från kamerorna. Termen bildkoordinat beskrivs i figur 4.



*Figur 4: Bildkoordinaterna i x- och y-led*

Bildkoordinaterna i x-led kan vara antingen positiva eller negativa beroende på deras position gentemot ett nolläge. För att senare möjliggöra uträkningar av vinklar som anger en riktning var objektet befinner sig relaterat till kameran måste nolläget vara beläget mitt i bilden så att pixlarna på ena sidan av nolläget är negativa och pixlarna på den andra sidan är positiva. Nolläget bestäms genom att dela antalet pixlar i x-led i två. Detta leder till att pixlarna på ena sidan av nolläget går från -1 till -n medan pixlarna på andra sidan nolläget går från +1 till +n var n är den sista pixeln i bilden. I själva positionsbestämningen leder detta till negativa och positiva vinklar då varje specifik x-koordinat leder till en unik vinkel. Relationen mellan negativa och positiva x-koordinater i en bild visas i figur 5.





*Figur 5: Negativa och positiva x-koordinater i en bild.*

## 2.2 Testtillfället för bestämning av bildkoordinater

Två algoritmer för ansiktsdetektion testades i detta arbete, nämligen Viola-Jones algoritmen (Viola & Jones, 2001) och en Haar-kaskad (Wikipedia, 2010a). Mjukvarorna som användes är FaceL för Viola-Jones algoritmen (Bolme et al., 2009) och Harpia projektet för Haar-kaskaden (S2i, 2007). Båda mjukvarorna är Open Source licensierat material som man hittar på Internet (Sourceforge, 2010a & b) För testet med detektion av cirklar användes det en Hough transform-algoritm som Harpia projektet erbjöd (S2i, 2007).

Kamerorna som användes till testandet var vanliga webbkameror av märkena Acer OrbiCam och TerraTec X2 med 1,3 megapixels CMOS sensorer. Uppdateringshastigheten på dessa kameror var 10 bilder i sekunden, vilket ledde till en baudrat på 100 ms.

Programmet FaceL kördes på en bärbar dator med Windows XP operativsystem och en 1,8 GHz processor. Mjukvaran Harpia kördes på en bordsdator med Linux Ubuntu som operativsystem och en 3,2 GHz processor.

På grund av de mjukvaror som fanns tillhanda kunde Viola-Jones algoritmen och Haar-kaskaden endast testas på ansikten och inga andra objekt. Harpia projektet erbjöd en möjlighet till cirkeldetektion och det var med en Hough transform.

Då koderna som Harpia projektet bidrog till inte i sig själv hade en funktion som mätte tiden för detektion av objektet i fråga måste koderna modifieras en aning. Ändringen bestod av några rader kod som gav CPU-tiden för detektionerna.

För att testa skillnaden mellan Viola-Jones algoritmen och Haar-kaskaden mättes tiden det tog för vardera algoritmen att hitta ett ansikte i bilden. Tio personer på en vit bakgrund användes som testgrupp. Dessa personer var på måfå utvalda frivilliga personer av varierande ålder och ursprung. Tre av testpersonerna var unga kvinnor, fem var unga män och två var äldre män. Två av personerna hade glasögon på sig under testet. Ansiktena var placerade mitt i bilden på båda kamerorna på ett avstånd av ca 0,5 m.

För att testa Hough transformens snabbhet att hitta cirklar i bilden användes tre bollar som testobjekt. Två av bollarna var mindre med diameter på 7,32 cm och en var större med en diameter på 11,9 cm. Den ena lilla bollen var vit, den andra röd och den stora bollen var grön. Som bakgrundsfärg användes vit, röd, brun, svart, grön, grå och blå-vit rutig.

Bollar valdes till testobjekt för cirkeldetektionen då deras form och utseende bildar en cirkel och är lika i alla riktningar och positioner. Detta argument utgör grunden till en lösning för problemet med att hitta en lämplig identifierande egenskap för hästarna i ett travlopp. Alternativet att använda olivfärgade bollar fästa på hästarna kunde vara en praktisk lösning även om den inte är väldigt estetisk.

## 2.3 Testresultat för bestämning av bildkoordinater

Vid ansiktsigenkänningen av tio på måfå valda personer mättes tiden det tog för mjukvaran att hitta ett ansikte i bilden. Kamerorna ställdes in så att ansiktet var ungefär mitt i bilden.

Mjukvaran som använde Viola-Jones algoritmen hittade ansikten i bilden som snabbast på 78 ms även om den allmännaste tiden för detektering var 94 ms. Den andra mjukvaran som använde en Haar-kaskad hittade ett ansikte på 390 ms medan de andra ansiktena i testgruppen tog 400-430 ms.

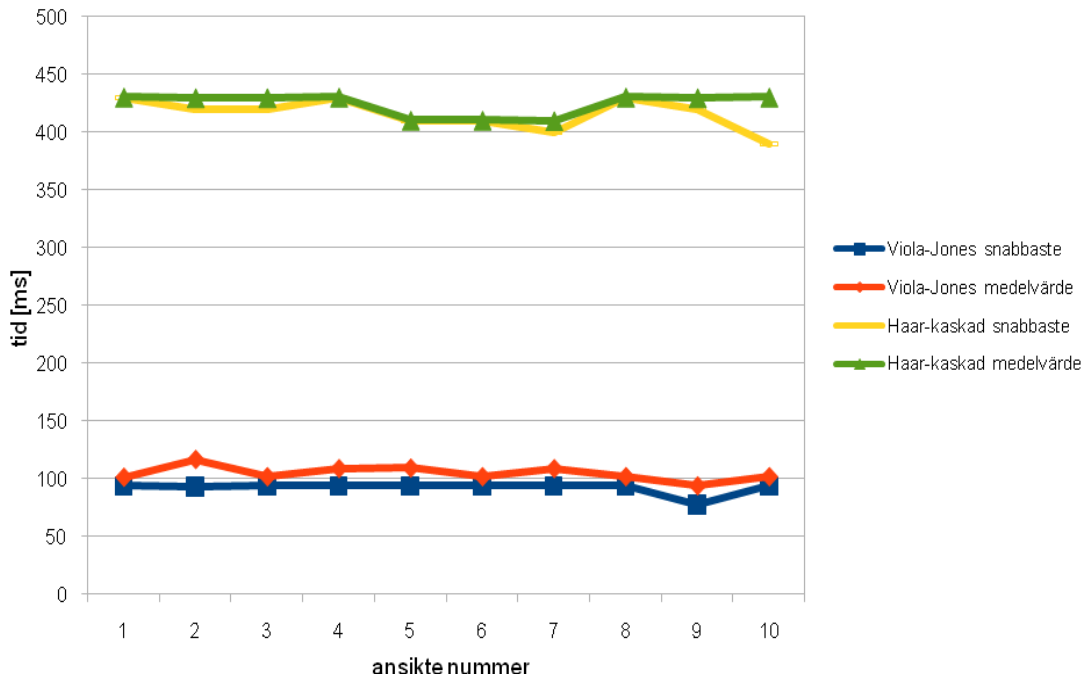
Testerna påvisar att Viola-Jones algoritmen är snabbare än en enkel Haar-kaskad. Enligt dessa test är Viola-Jones algoritmen 4,6 gånger snabbare än Haar-kaskaden. Om den bärbara datorn som testade Viola-Jones algoritmen skulle ha haft en lika snabb processor som bordsdatorn som exekverade Haar-kaskaden kunde man anta att detektionen av ett ansikte borde ha skett på 53 ms. Då skulle Viola-Jones algoritmen var 8,1 gånger snabbare än en enkel Haar-kaskad.

De snabbaste tiderna för vardera algoritmen och för varje ansikte samt medelvärdet mellan den snabbaste och den långsammaste tiden för varje detektion är presenterade i tabell 1. Värdena inom parentes för Viola-Jones algoritmen är den antagna tiden det skulle ta att detektera ett ansikte med samma processor som för Haar-kaskaden, det vill säga med en 3,2 GHz processor istället för en 1,8 GHz processor. Grafen i figur 6 förevisar den snabbaste detektionen för varje ansikte.

*Tabell 1: Jämförelse över ansiktsgenkänning mellan Viola-Jones algoritmen och Haar-kaskaden. Tiderna inom parentes vid Viola-Jones algoritmen anger den antagna tiden för detektionen då datorns processor varit den samma som för Haar-kaskaden, det vill säga 3,2 GHz.*

<b>Ansikte nr.</b>	<b>Viola-Jones snabbaste [ms]</b>	<b>Viola-Jones medelvärde [ms]</b>	<b>Haar-kaskad snabbaste [ms]</b>	<b>Haar-kaskad medelvärde [ms]</b>
1	94 (52,2)	101,5	430	430,5
2	93 (51,6)	117	420	430
3	94 (52,2)	102	420	430
4	94 (52,2)	109	430	430,5
5	94 (52,2)	109,5	410	410,5
6	94 (52,2)	102	410	410,5
7	94 (52,2)	109	400	410
8	94 (52,2)	102	430	430,5
9	78 (43,3)	94	420	430
10	94 (52,2)	102	390	430,5

## Jämförelse mellan Viola-Jones och Haar-kaskad



Figur 6: Jämförelse mellan Viola-Jones algoritmen och Haar-kaskaden.

Testet med koden som detekterar cirklar utfördes med TerraTec webbkameran på en bordsdator med Linux Ubuntu operativsystemet. Bollarna placerades framför bakgrunden så att de var belägna ungefär mitt i kamerans bild. Avståndet från kameran till bollarna var ca. 30 cm. Det som resultaten visar är att Hough transformen kan utföra en cirkel detektion ungefär tio gånger snabbare än en Haar-kaskad kan utföra en ansiktsdetektion och 0,89 gånger snabbare än en Viola-Jones algoritmen kan antas utföra en ansiktsdetektion då den exekveras av en 3,2 GHz processor.

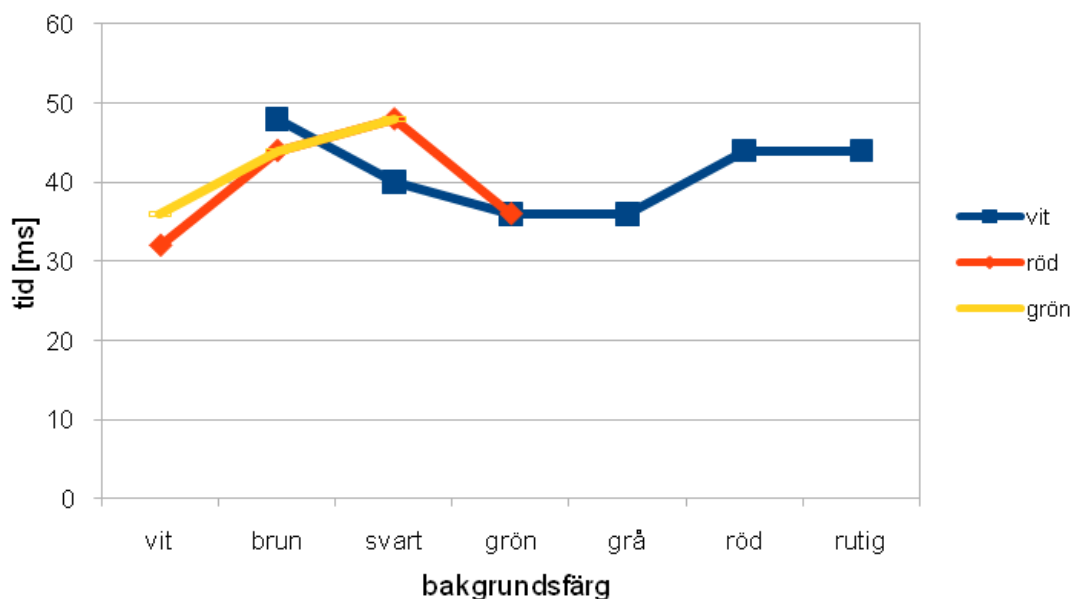
Problemet med denna typ av detektion är att en boll av samma färg som bakgrunden inte detekteras. Dessutom visar resultaten att bollar och bakgrunder som är kontrastfärg till varandra detekteras inte. Även om den röda bollen på den gröna bakgrunden emellanåt detekterades kunde inte en stabil och kontinuerlig detektering uppnås. Den vita bollen var det objektet som säkrast kunde detekteras förutom om den var placerad på en vit bakgrund.

Den vita bakgrunden var enligt testerna lättast för programmet att exekvera medan den bruna och den svarta bakgrunden var svåra för mjukvaran att handskas med. Resultaten från testerna är framförda i tabell 2 och i diagrammet i figur 7.

*Tabell 2: Testresultaten av cirkeldetektionen med en Hough transform.*

<b>Bakgrunds färg</b>	<b>Vit, liten boll snabbaste</b>	<b>Vit, liten boll medelvärde</b>	<b>Röd, liten boll snabbaste</b>	<b>Röd, liten boll medelvärde</b>	<b>Grön, stor boll snabbaste</b>	<b>Grön, stor boll medelvärde</b>
vit	-	-	32	34	36	38
brun	48	74	44	70	44	74
röd	44	46	-	-	-	-
svart	40	44	48	78/-	48	80
grön	36	38	36	38/-	-	-
grå	36	38	-	-	-	-
rutig	44	44	-	-	-	-

## Cirkeldetektion med olika bakgrunder



Figur 7: Cirkeldetektion med olika bakgrunder.

Testerna visar att man kan utföra en detektion av en cirkel på 32 ms som snabbast med en Hough transform vid goda förhållanden. Medelvärde för alla detektioner i testet för cirkeldetektion är 47,4 ms.

Testen med ansiktsdetektionen påvisar också det att Viola-Jones algoritmen är snabbare än traditionella algoritmer som Haar-kaskaden. Det som förblir oklart är om Viola-Jones algoritmen kan detektera cirklar snabbare och säkrare än Hough transformen.

## 2.4 Slutsats för detektion och identifiering

Det som påvisas i testerna som utfördes var att Viola-Jones algoritmen är åtminstone 4,6 gånger snabbare än en Haar-kaskad. Enligt forskarna Paul Viola och Michael Jones skall deras algoritm vara snabbare än någon annan känd ansiktsdetektionsalgoritm (Viola&Jones, 2001). Hur snabbt Viola-Jones algoritmen kan detektera cirklar borde testas för att försäkra sig om att den metoden är snabbare än Hough transformen som användes i testerna för detta projekt. Det finns dock skäl att anta att en cirkeldetektion med Viola-Jones algoritmen kunde ske en hel del snabbare än men en Hough transform.

Det som framkom i testerna med cirkeldetektionen var att en ljusare cirkel detekteras lättare på en mörk bakgrund än en färgad eller mörk cirkel det vill säga en viss kontrast krävs för att en cirkel skall detekteras. Då bakgrunden i det slutliga ändamålet kommer att vara mer mörk och brokig än ljus och enfärgad vore det fördelsaktigt om de möjliga identifierande bollarna på hästarna skulle vara ljusa till sin färg. Ett alternativ till ljusa bollar kunde vara att invertera färgerna i bilden för att underlätta identifieringen av hästarna. Då skulle en mörk boll vara ljus i bilden från kameran och på så sätt vara lättare att urskilja.

Då medelvärdet för cirkeldetektionen med en Hough-transform visade sig vara det snabbaste medelvärdet för detektioner i de test som utfördes kunde man anta att tiden  $\Delta t_2$  i figur 2 skulle vara minst 47,4 ms. Denna tid skulle leda till att en häst som rör sig med 16 m/s skulle hinna röra sig 760 mm under den tid som krävs för detektering och identifiering av hästarna. Vid detta skede då tiden  $\Delta t_3$  inte ännu är känd är fördröjningen från realtid 87,4 ms. Detta betyder att en häst med hastigheten 16 m/s skulle hinna röra sig 1398,4 mm innan föregående position har bestämts.



## 3 POSITIONSBESTÄMNING MED TRIANGULERING

### 3.1 Inledning

Då bildkoordinaterna för objektet är bestämda i alla kameror kan en positionering i världskoordinater ske. Till detta krävs det bilder från två eller flera kameror som filmat samma objekt från olika ställen. Kamerorna måste vara stationära i alla led för att en position skall kunna beräknas. Om zoomande objektiv önskas användas måste information om brännviddens längd vara tillhanda då varje bild tas. Om detta inte är möjligt måste zoomen fixeras till ett läge med en känd längd på brännvidden.

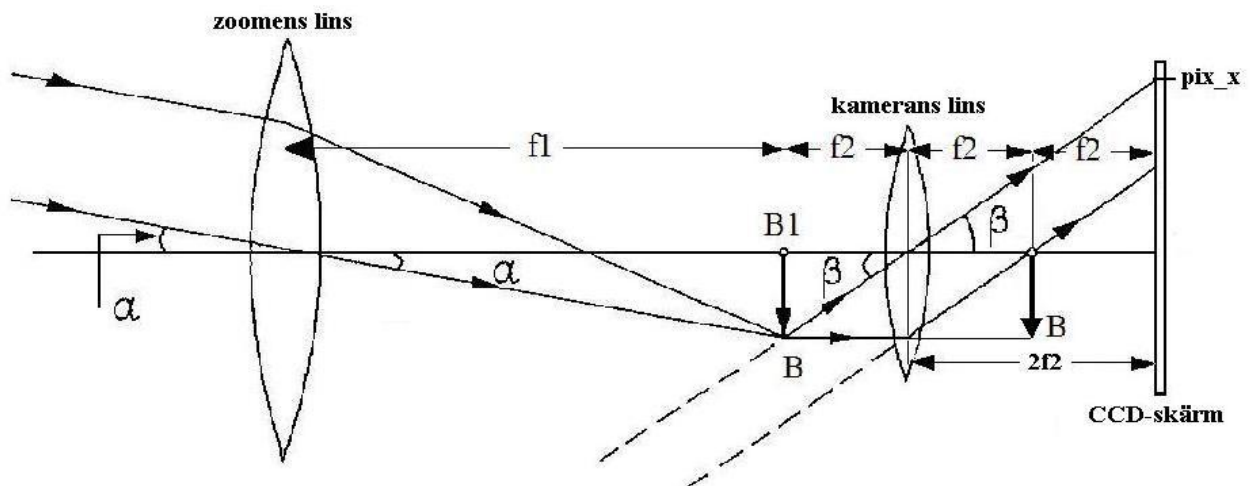
Tekniken bakom positionsbestämningen som tillämpas i detta arbete är triangulering som använder sig av storleken på vinklarna och längderna på kateterna i trianglar för att bestämma en viss position. Matematiken bakom de uträkningar som presenteras nedan är trigonometriska funktioner som man hittar i alla matematik- och tabellböcker.

### 3.2 Teorin bakom triangulering

Positionen i världskoordinater bestäms genom att räkna ut de vinklar och sträckor som sammanbinder kamerorna med objektet till en triangel. Positioneringen utgår från bildkoordinaterna i x-led från varje kamera. Med x-koordinaten och de förhållanden som råder i kamerans optiska system då bilden togs kan man räkna ut vinkeln  $\alpha$  för varje kamera. Vinkeln  $\alpha$  beskriver vidden på kamerans bildområde och ger en riktning från kameran mot objektet vilket leder till att man kan räkna ut de sträckor som sammanbinder kamerorna och objektet. Vinkeln  $\alpha$

påverkas av den zoomande linsens fokallängd som förändras då zoomens lins förflyttas närmare eller längre bort från kamerans lins beroende på hur mycket man vill förstora bilden. Brännvidden på de zoomande objektiven som kamerorna på Vermo travbana är utrustade med varierar mellan 9 och 153 mm och brännvidden på kamerornas linser är 15,24 mm.

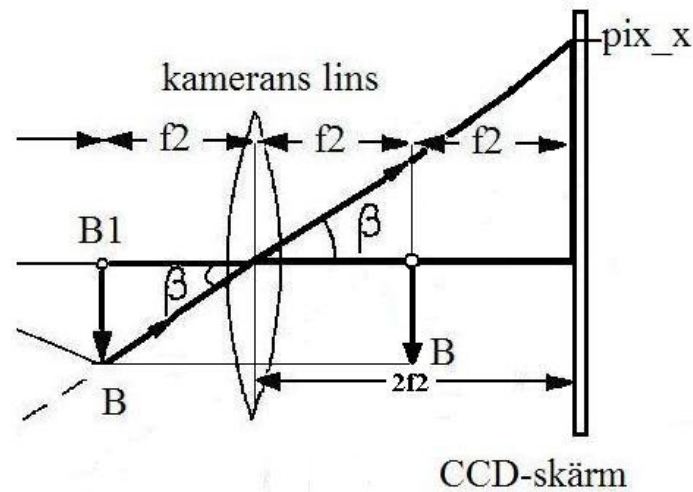
Kamerans interna optiska system utgör de fysikaliska utgångspunkterna för den matematiska funktionen som gör det möjligt att räkna ut vinkeln  $\alpha$ . En förenklad bild av det optiska systemet förevisas i figur 8.



Figur 8: De fysikaliska förhållandena mellan kamerans- och zoomens linser i en kamera.

I figur 8 bildar en inkommande stråle en bildkoordinat  $x$  på CCD-skärmen genom att färdas först genom den zoomande linsen och sedan genom kamerans lins. De vinklar som strålen bildar med den optiska axeln då den passerar linserna är  $\alpha$  respektive  $\beta$ . Vinkeln  $\beta$  beskriver vidden på bilden utan en zoomande lins medan vinkeln  $\alpha$  är vidden på bilden då en zoomande lins är i användning. Då den zoomande linsen tillsammans med kamerans lins bildar det optiska systemet och båda linsernas egenskaper påverkar kamerans bildvinkel är vinkeln  $\alpha$  den väsentliga vinkeln som används vid beräkning av ett objekts position.

Ekvationen för vinkeln  $\alpha$  är härledd från två enskilda ekvationer, en för vinkeln  $\beta$  och en för vinkeln  $\alpha$  i figur 8. Förhållandena som leder till ekvationen för vinkeln  $\beta$  är förevisade i figur 9.



Figur 9: Fysikaliska förhållandena för vinkeln  $\beta$ .

Vinkeln  $\beta$  som bildas vid kamerans lins mellan den inkommande strålen och den optiska axeln bestäms av fokallängden för den linsen och den x-koordinat som bildas där var stålen träffar skärmen. Avståndet från linsen till CCD-skärmen är två gånger fokallängden  $f_2$  och avståndet från linsen till punkten B är fokallängden  $f_2$ . Dessa förhållanden leder till två ekvationer för vinkeln  $\beta$ :

$$\tan \beta = \text{pix\_x}/2f_2 \quad (1)$$

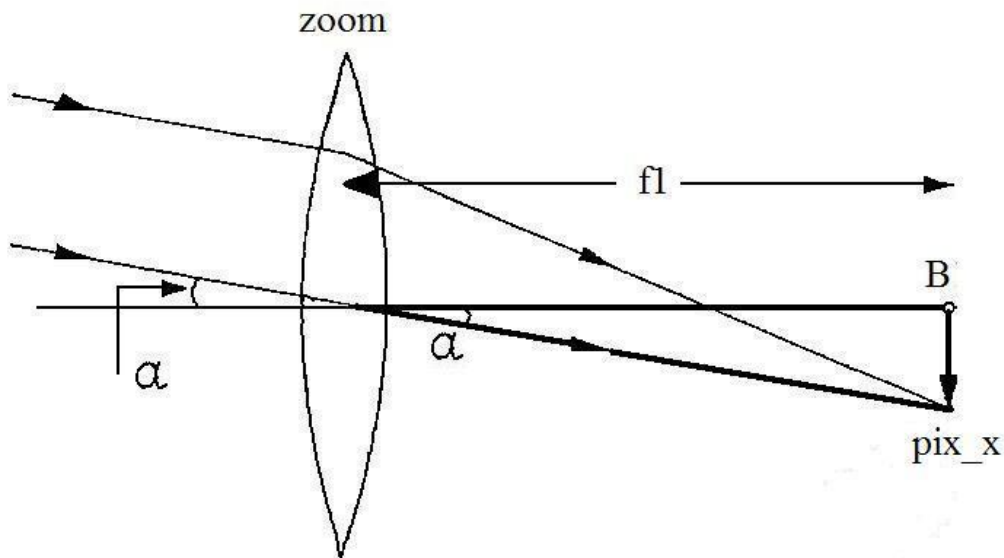
$$\tan \beta = B/f_2 \quad (2)$$

var x-koordinaten och B utgör den motstående kateten medan fokallängden utgör den närliggande kateten enligt figur 9. Då x-koordinaten och B båda utgör den motstående kateten bestäms förhållandet mellan dem enligt

$$\text{Pix}_x/2f_2 = B/f_2 \quad \longrightarrow \quad B = \text{Pix}_x/2 \quad (3)$$

Förhållandet mellan  $\text{Pix}_x$  och  $B$  kommer att vara av stor vikt vid härledningen av ekvationen för vinkeln  $\alpha$  då de utgör den motstående kateten i den tangensfunktion som kommer att användas.

Vinkeln  $\alpha$  härleds på samma sätt som vinkeln  $\beta$  från de fysikaliska förhållandena i kamerans optiska system enligt figur 10.



Figur 10: De fysikaliska förhållandena för uträkningarna av vinkel  $\alpha$ .

För vinkeln  $\alpha$  gäller ekvationen

$$\tan \alpha = B/f_1 \quad (4)$$

var  $B$  utgör den motstående kateten och har ett värde hälften så stort som  $\text{Pix}_x$  enligt ekvation (3) medan  $f_1$  utgör den närliggande kateten och har ett värde av fokallängden  $f_1$  för den zoomande linsen.

Då både kamerans och zoomens linser är belägna i samma optiska system enligt figur 8 härleds vinkeln  $\alpha$  genom att kombinera ekvationerna för både vinkeln  $\alpha$  och  $\beta$ . Från detta får man två ekvationer för den riktiga vinkeln  $\alpha$  enligt

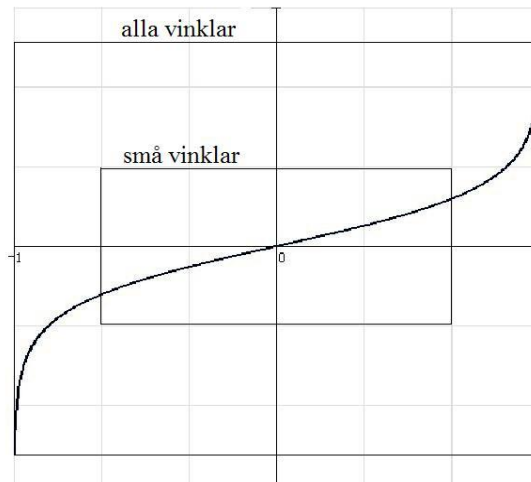
$$\alpha_{\text{linjär}} = \text{Pix}_x/2f1 \quad (5)$$

var  $\text{Pix}_x$  är x-koordinaten i bilden och  $f1$  är fokallängden för den zoomande linsen. Denna ekvation är en linjär regression och lämpar sig för små värden på vinkeln  $\alpha$ . En allmän formel som lämpar sig för alla värden på vinkeln  $\alpha$  är

$$\alpha_{\text{allmän}} = (180/\pi) * \arctan(\text{Pix}_x/2f1) \quad (6)$$

var  $(180/\pi)$  är en förändring från radianer till grader,  $\text{Pix}_x$  är x-koordinaten i bilden och  $f1$  är fokallängden för den zoomande linsen.

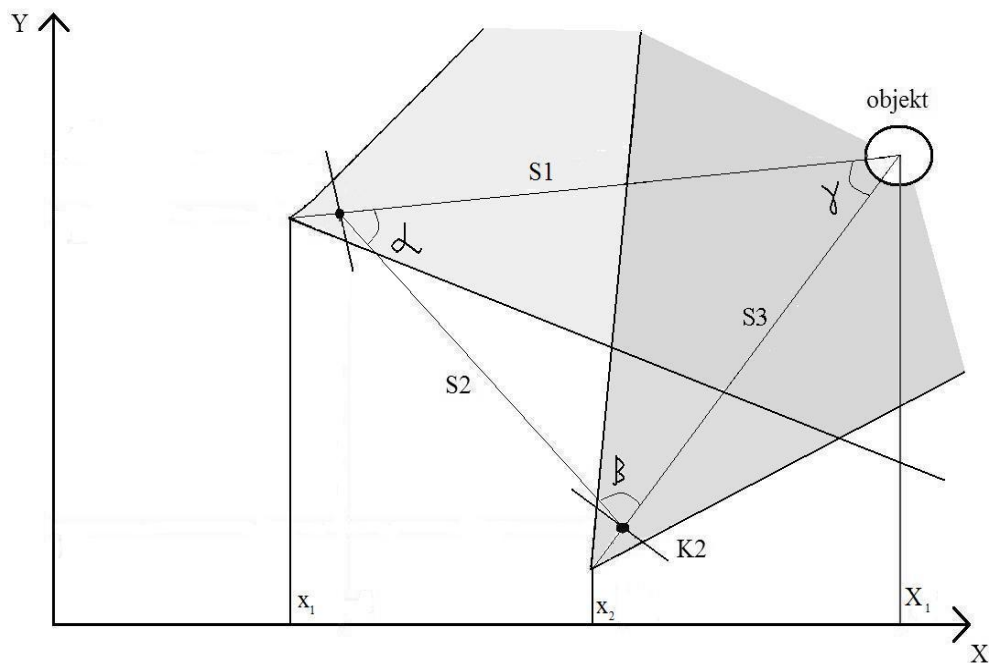
Lämpligheten av ekvationerna (5) och (6) för olika värden på vinkeln  $\alpha$  har sin grund i tangensfunktionen som är nästintill linjär då den närmar sig origo men olinjär vid stora värden. Detta betyder att den linjära ekvationen gäller för små värden på vinkeln  $\alpha$  medan den allmänna ekvationen gäller för alla värden på vinkeln  $\alpha$  då tangensfunktionen har tagits i beaktande i ekvationen. Detta förhållande är beskrivet i figur 11.



*Figur 11: Tangens funktionen och dess relevans vid vinkeln  $\alpha$  storlek.*

Linsernas egenskaper att bryta de inkommande strålarna och x-koordinatens position med nollläget i mitten av bilden enligt figur 5 leder till att förhållandet mellan vinklarna  $\alpha$  och  $\beta$  är inverterat. Detta betyder att om vinkeln  $\beta$  är positiv kommer vinkeln  $\alpha$  att vara negativ och vice versa.

Då vinkeln  $\alpha$  är bestämd för varje kamera och sträckan mellan de stationära kamerorna är känd kan man räkna ut sträckorna mellan kamerorna och objektet med sinussatsen. För att förevisa en triangulering av ett objekt presenteras det i figur 13 en situation med två kameror.



Figur 12: Triangulering med två kameror.

Kamerorna K1 och K2 filmar ett objekt. Kamerorna och objektet bildar hörnen i den triangel som sammanbinds av sträckorna S1, S2 och S3 med vinklarna  $\alpha$ ,  $\beta$  och  $\gamma$ . Kamerornas hörn har sina specifika bildkoordinater  $x_1$  och  $x_2$  medan objektet har sin specifika världskoordinat  $X_1$ .

De olikfärgade områdena visar de områden kamerorna filmar. De två ljusare områdena beskriver det område som endast den ena kameran kan avbilda medan det mörkare området är det område som båda kamerorna kan filma och då är en positionsbestämning genom triangulering möjlig. Detta leder också till att objektet måste vara på ett visst avstånd från kamerorna för att den skall synas i bilderna från alla kameror.

Vinklarna  $\alpha$  och  $\beta$  är de uträknade vinklarna  $\alpha$  för varje kamera som presenterades i figur 8 och deras storlekar bestämdes av antingen ekvation (5) eller (6). Förhållandet mellan hörnen på triangeln samt kateterna i den är enligt sinussatsen

$$S1/\sin \beta = S2/\sin \gamma = S3/\sin \alpha \quad (7)$$

var sträckan  $S2$  och vinklarna  $\alpha$  och  $\beta$  är kända. Den tredje vinkeln  $\gamma$  kan bestämmas genom triangelns vinkelsumma vilket betyder att vinkeln  $\gamma$  är

$$\gamma = 180 - (\alpha + \beta) \quad (8)$$

Detta leder till att sträckorna  $S1$  och  $S3$  kan räknas ut enligt

$$S1 = S2 * [\sin \beta / \sin(180 - (\alpha + \beta))] \quad (9)$$

$$S3 = S2 * [\sin \alpha / \sin(180 - (\alpha + \beta))] \quad (10)$$

Positionen i världskoordinater för objektet fås då alla de uträknade sträckorna och vinklarna sammanfaller och bildar en triangel mellan kamerorna och objektet.

### 3.3 Testtillfälle för triangulering

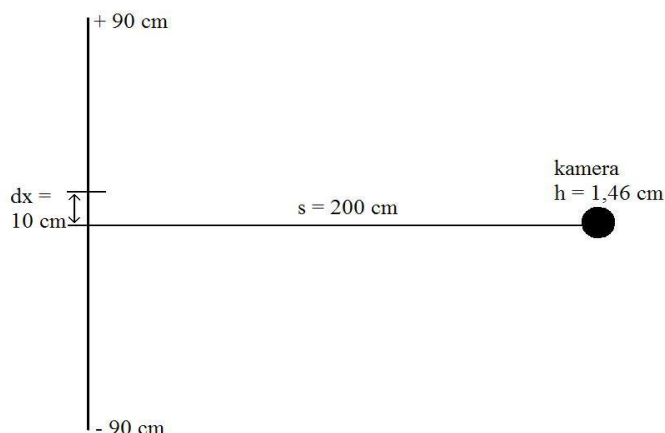
Testet som presenteras nedan utfördes för att bevisa teorierna bakom trianguleringen samt för att jämföra de två ekvationerna för vinkeln  $\alpha$ . Testet utfördes med en Haar-kaskad ansiktsdetektionsalgoritm då den mjukvaran hade egenskapen att ge ut bildkoordinaterna för det detekterade ansiktet. De förhållanden som gällde vid testtillfället är uppräknade i tabell 3.

Tabell 3: Förhållandena vid testet för triangulering.

Kamera modell	Terratec X2
Bilder per sekund	10 fps
Antal pixlar i bilden	640x480
Höjd på kameran	146 cm
Avstånd kamera-ansikte	200 cm
Avstån mellan mätningar (dx)	10 cm
Strörsta/minsta dx	+/- 90 cm
Detektionsalgoritm	Haar-kaskad
Processorhastighet	3,2 GHz
Operativsystem	Linux Ubuntu

Kameran på stativet ställdes in så att den var på samma höjd som ansiktet, det vill säga på 1,46 m höjd. Ansiktet befann sig på ett konstant avstånd på 2,00 m från kameran. Testsituationen är avbildad i figur 12.





Figur13: Testsituation för triangulering.

Med denna höjd på kameran och på detta avstånd från den befann sig ansiktet nästan mitt i bilden med bildkoordinaterna 305x289 i en bild på 640x480 pixlar. Detta läge utgjorde bildens nolläge för de uträkningar som presenteras nedan.

Ansiktet förflyttades 10 cm åt gången först till höger och senare till vänster. Vid  $dx = 100$  cm syntes inte hela ansiktet mer i bilden och kunde på såvis inte detekteras.

Varje förflyttning  $dx$  åt endera sidan åstadkom en specifik bildkoordinat i x-led. En förflyttning åt vänster ledde till en större bildkoordinat än nolläget vilket gav en positiv position i bilden medan en förflyttning till höger orsakade en mindre bildkoordinat i x-led än nolläget vilket gav en negativ position i bilden.

På grund av att nolläget är mitt i bilden och inte i nedre vänstra hörnet måste en korrigering i x-koordinaterna ske för att vinkeln  $\alpha$  skall bli korrekt. Nolläget i detta test utgjordes av pixel 305 i x-led vilket betydde att koordinaterna för varje enskilt  $dx$  subtraherades från detta nolläge. På så sätt fick man en x-axel som räckte sig från -260 till +260 med ett nolläge mitt i bilden.

Eftersom webbkameran inte hade en zoomande lins beräknades vidden på kamerans bildområde enligt figur 9 var vinkeln  $\beta$  är den enda vinkeln som påverkar vidden på bilden. Detta leder till att vinkeln  $\beta$  benämns som  $\alpha$  i dessa uträkningar.

Vinkeln  $\alpha$  räknades ut med både den linjära ekvationen (4) och den allmänna ekvationen (3). För att försäkra sig om ekvationerna (3) och (4) riktighet räknades det ut de teoretiska värdena på vinkeln  $\alpha$  vilka borde ha orsakats av de rådande förhållandena. De teoretiska värdena på vinkeln  $\alpha$  bestämdes med ekvationen

$$\alpha_{\text{teoretisk}} = \arctan(dx/s) \quad (11)$$

var  $dx$  är förflyttningen i sidled och  $s$  är sträckan från kameran till ansiktet, båda variablerna är uppritade i figur 14.

Då den linjära ekvationen lämpar sig för små värden på vinkeln  $\alpha$  vilket gällde för dessa testresultat, bestämdes linjens lutning för värdena från testtillfället. Då fick man ekvationen

$$\alpha_{\text{linjär}} = 0,0934 * Pix\_x \quad (12)$$

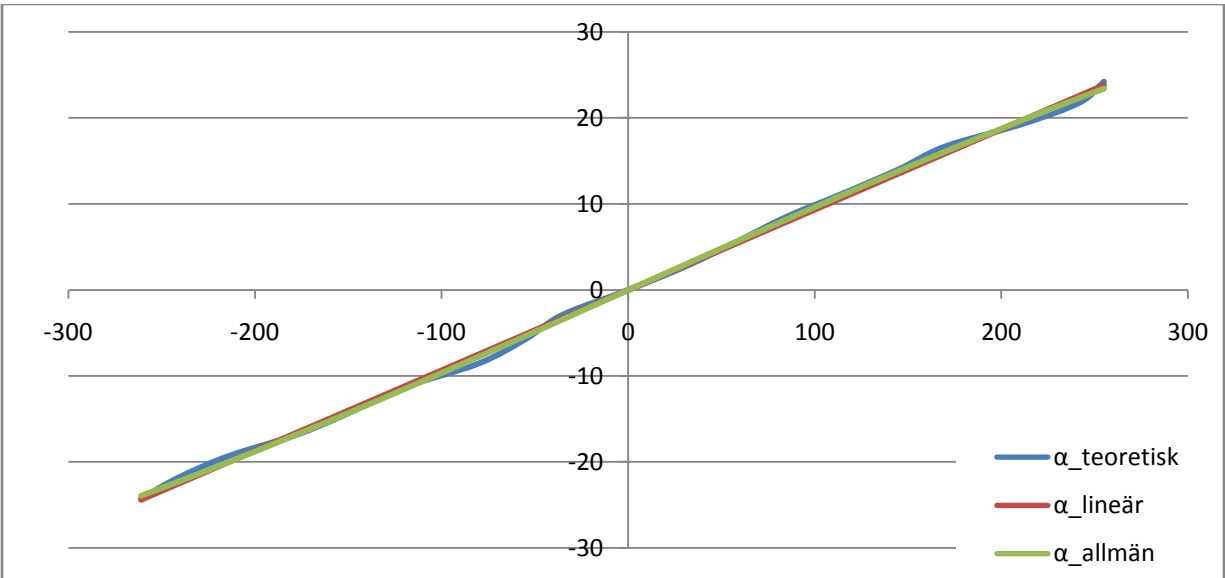
var konstanten 0,0934 är linjens lutning och  $Pix\_x$  är den korrigerade bildkoordinaten i x-led.

För de förhållanden som gällde vid testtillfället blev den förenklade allmänna ekvationen

$$\alpha_{\text{allmän}} = 180/\pi * \arctan(Pix\_x * 0,0017) \quad (13)$$

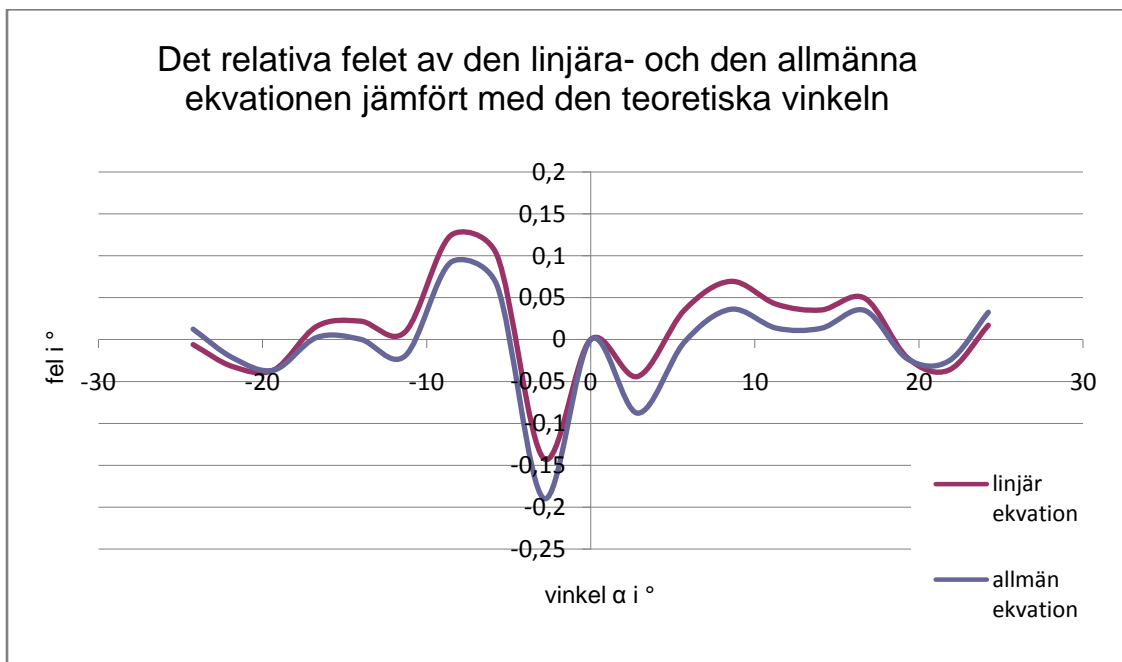
var  $180/\pi$  står för ändring av radianer till grader,  $Pix\_x$  är den korrigerade x-koordinaten i bilden 0,0017 är en konstant för fokallängden  $f_1$ .

För att undersöka om dessa ekvationer sammanfaller med de teoretiska värdena för vinkeln  $\alpha$  ritades resultaten av alla tre ekvationerna upp i figur 15.

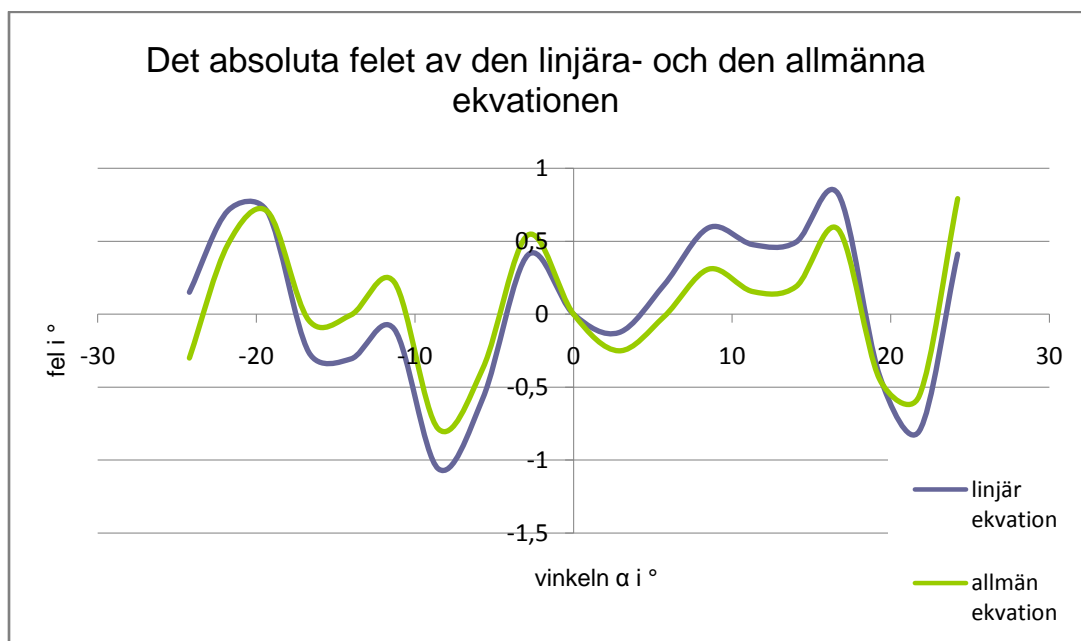


Figur 14: Jämförelse mellan de tre ekvationerna för uträkning av vinkeln  $\alpha$ .

Av figuren kan man dra slutsatsen att både den linjära ekvationen och den allmänna ekvationen ger ett sanningsenligt värde för vinkeln  $\alpha$  även om ett litet fel för vardera ekvationen kan urskiljas. Det relativa felet för båda ekvationerna är presenterade i figur 15 och det absoluta felet är förevisat i figur 16. I tabell 4 är de maximala absoluta och relativa felen för ekvationerna presenterade.



Figur 15: Det relativa felet av den linjära och den allmänna ekvationen jämfört med den teoretiska vinkeln.



Figur 16: Det absoluta felet av den linjära och den allmänna ekvationen.

Tabell 4: Den linjära och den allmänna ekvationens relativa och absoluta fel.

<b>Ekvation</b>	<b>Relativt fel max</b>	<b>Relativt fel medelvärde</b>	<b>Absolut fel max</b>	<b>Absolut fel medelvärde</b>
Linjär ekvation	0,12	0,01	0,82	0,68
Allmän ekvation	0,092	-0,005	0,79	0,06

Med dessa resultat som grund kan man konstatera att den allmänna ekvationen har ett mindre fel än den linjära ekvationen och är därför mer lämpad för triangulering av ett objekt på ett målområde. Detta leder till att man kan använda en ekvation för uträkningar på vinkeln  $\alpha$  oberoende av dess storlek.

### 3.4 Slutsats för positionering med triangulering

Till en positionering i två dimensioner med ett antal kameror krävs det en ekvation för att räkna ut positionen i världskoordinater för ett objekt på ett målområde, nämligen ekvation (6).

Denna ekvation behöver 20 instruktioner för att en processor skall kunna exekvera den. För att exekvera en vanlig addition krävs det två instruktioner och för att exekvera en multiplikation krävs det fyra instruktioner. Om en vanlig dator med en processor på t.ex. 1 GHz skall köra en addition kommer det att ta 2 ns för mjukvaran att utföra uppgiften. För att exekvera 20 instruktioner med samma dator tar det 20 ns och för att utföra 1000 instruktioner skulle det ta 1  $\mu$ s.

Genom detta resonemang kan man anta att tiden  $\Delta t_3$  i figur 3 kommer att vara minimal och inte försämra den fördröjning från realtiden som orsakas av kamerans egenskaper och identifieringen av objektet i bilden.

Enligt det test som utfördes för att bevisa teorierna om triangulering kommer det absoluta felet att vara  $0,79^\circ$  eller jämförelsevis 0,79 % med den allmänna ekvationen som visade sig var den noggrannare. Ett fel av denna storlek kommer att orsaka ett grovt positioneringsfel. Felet som ekvationen orsakar kan minskas genom att öka mängden pixlar i bilden. Mängden pixlar som skulle krävas för att en position kunde bestämmas med en centimeters noggrannhet härleds nedan.

Distansen mellan en kamera och objektet kan bestämmas genom

$$S_a = S_0(\sin \alpha / \sin \gamma) \quad (14)$$

var  $S_a$  är sträckan mellan kameran och objektet,  $S_0$  är sträckan mellan kamerorna och  $\sin \gamma$  är sinus värdet för vinkeln  $\gamma$  vid objektet som bestäms genom  $180^\circ - (\alpha + \beta)$ . För att få det relativa felet i distansen som orsakas av den allmänna ekvationen deriverar man ekvation (14) enligt

$$dS_a/d\alpha = \cos \alpha / \sin \gamma \quad (15)$$

vilket leder till att man får felet i distansen som  $\Delta S_a$  enligt

$$\Delta S_a = \sqrt{\left(\frac{dS_a}{d\alpha}\right)^2 * (\Delta\alpha)^2} = |dS_a/d\alpha| * |\Delta\alpha| \quad (16)$$

Detta leder till att felet  $\Delta S_a$  jämfört med den verkliga distansen  $S_a$  är

$$\frac{\Delta S_a}{S_a} = \left| \frac{\cos \alpha / \sin \gamma}{\sin \alpha / \sin \gamma} \right| * \Delta\alpha = \cot \alpha * \Delta\alpha \quad (17)$$

var  $\cot \alpha$  är cotangens funktionen av vinkeln  $\alpha$  och  $\Delta\alpha$  är det relativa felet grader i vinkeln  $\alpha$ .

I testresultaten från det test som utfördes förekommer det i medeltal ett fel på 0,209 m i distansen mellan kameran och ansiktet. Då kravet på det tillåtna felet är 0,01 m betyder det att mängden pixlar i bilden borde vara 20,9 gånger fler vilket är 13376 pixlar i x-led. Med det bildformat på 960x540 pixlar som används på Vermo travbana i nuläget skulle felet i distansen mellan en kamera och objektet vara 0,139 m.

Då ett bildformat med 13376 pixlar i x-led inte finns, kan en noggrannhet på 0,01 m uppnås genom att använda tillräckligt med zoom. I testet som utfördes skulle en sådan noggrannhet kräva en 21 gånger förstörande zoom på webbkameran. På Vermo travbana där bildformatet är en aning större, vilket minskar felet till 0,139 m, skulle det krävas en förstoring på 14 gånger av bilden för att uppnå den önskade noggrannheten. Detta går att uppnå med de kameror som redan är i användning där då de har en zoom som förstörar bilden upp till 17 gånger.

I det tänkta systemet räcker det inte med endast två kameror för positioneringen av objektet på grund av målområdets storlek och form. Fler än två kameror ökar säkerheten av positionsbestämningen då fler sammanband bör uppfyllas för att en position skall hittas.

## 4 DISKUSSION

Den totala tiden  $T_{tot}$  för processen i figur 2 utgörs av  $\Delta t_1$  vilket har konstaterats vara 40 ms på grund av det videoformat som används,  $\Delta t_2$  som genom testning kunde fastställas till 47,4 ms

som ett medelvärde av cirkeldetektionerna med en Hough transform samt  $\Delta t_3$  vilken kommer att vara minimal till sin omfattning och inte påverka den totala tiden. Detta leder till att  $T_{\text{tot}} = 87,4$  ms vilket betyder att ett positionsbestämningssystem med dessa förutsättningar kommer att vara minst 87,4 ms fördröjt från den absoluta realtiden. Den totala tiden kan förbättras genom att använda ett snabbare videoformat och utveckla en snabbare detektions- och identifieringsalgoritm.

Felet i sträckorna som sammanbinder kamerorna med objektet kommer att vara 0,139 m då ingen zoom används i den bild på 960x540 pixlar som kamerorna på Vermo travbana har. För att felet i sträckorna skall vara 0,01 m krävs det en zoom som konstant förstörar bilden 14 gånger.

Den optimala placeringen för kamerorna är  $360^\circ$ /antalet kameror (Jusslin&Lundquist, 2004, s 7-8) vilket ger möjligheten till att använda tre av de fyra kameror som finns på Vermo travbana. Då takkamerorna är belägna parallellt med varandra, alltså i en vinkel på  $180^\circ$ , kompenserar de ut varandra så att det inte är lönsamt att använda båda takkamerorna till en positionsbestämning med triangulering. Då målområdet är relativt stort och har ett område i mitten som inte skall ingå i positionsbestämningen av objekten kan det behövas fler än tre kameror för att utföra positioneringar över hela målområdet.

För att identifiera hästarna på målområdet kunde en färgad boll fäst på hästen vara en användbar lösning även om den inte är väldigt estetisk. För att försnabba detektionen och identifieringen av hästarna borde man utveckla en cirkel- och färgdetekterande Viola-Jones algoritm. Med denna metod kunde man nå en mer precis identifiering och minska den totala fördröjningen.

Ett system uppbyggt med de teorier som har framförts i detta arbete skulle en kontinuerlig positionsbestämning i världskoordinater vara fullt möjlig. Det finns dock vissa praktiska problem som bör lösas om detta system skall praktiskt genomföras.

## KÄLLFÖRTECKNING

Bolme Davis S., Beveridge J. Ross, Draper Bruce A. 2009. FaceL, Facile Face Labeling. Colorado State University, 10s.

Jusslin Markus & Lundqvist Per. 2004. Kamerabaserad positionsbestämning av mark- och luftbrisader. Umeå Universitet, Institutionen för datavetenskap, 120s.

Panasonic. 2008. 3CCD Convertible Camera Pan-tilt Systems Compact Live switchers [www]. Hämtat 17.12.2009.

[http://panasonic.net/pss/broad/catalog/pdf/3CCD\\_ConvertibleCamera\\_WR.pdf](http://panasonic.net/pss/broad/catalog/pdf/3CCD_ConvertibleCamera_WR.pdf)

S2i- Industrial Intelligent Systems.2007. The Harpia project [www]. Hämtat 29.3.2010.  
<http://s2i.das.ufsc.br/harpia/en/home.html>

Sourceforge. 2010a. FaceL: Facile Face Labeling [www]. Hämtat 17.2.2010.

<http://sourceforge.net/projects/pyvision/files/FaceL/>

Sourceforge. 2010b. Project Harpia [www]. Hämtat 17.2.2010.

<http://sourceforge.net/projects/harpia/>



Suomen Hippos. 2009. Raviurheilu [www]. Hämtat 17.1.2010.

<http://www.hippos.fi/hippos/raviurheilu/index.php>

Viola Paul & Jones Michael. 2001. Robust Real-time Object Detection. Second International Workshop on Statistical and Computational Theories of Vision – Modeling, Learning, Computing and Sampling. Vancouver, Kanada, 25s.

Wikipedia. 2010a. Haar-like features [www]. Hämtat 17.2.2010.

[http://en.wikipedia.org/wiki/Haar-like\\_features](http://en.wikipedia.org/wiki/Haar-like_features)

Wikipedia. 2010b. Serial Digital Interface [www]. Hämtat 14.2.2010.

[http://en.wikipedia.org/wiki/Serial\\_Digital\\_Interface](http://en.wikipedia.org/wiki/Serial_Digital_Interface)