Visuell intelligens ger roboten rumskänsla

EXPERTARTIKEL

Algoritmerna och hårdvaran som skänker roboten syn

Av Brian Dipert, Embedded Vision Alliance, Yves Legrand, Freescale Semiconductor, Bruce Tannenbaum, MathWorks

Brian Dipert är chefredaktör för Embedded Vision Alliance och senior analytiker på BDTI (Berkeley Design Technology, Inc.) som erbjuder analys, rådgivning, och utveckling av inbyggda system och tillämpningar. Dessutom är han chefredaktör för InsideDSP, företagets webbnyhetsbrev om digital signalbehandling. Brian Dipert läste till elektroingenjör (B.S.) på Purdue University i West Lafayette, Illinois. Sina första yrkesaktiva år tillbringade han på Magnavox Electronics Systems i Fort Wayne. Han har också hunnit med åtta år på Intel i Folsom och 14 år på EDN Magazine.

Yves Legrand är global vertikal marknadschef för industriautomation och robotsystem på Freescale. Han kommer från Frankrike och har fördelat sitt karriärliv mellan Toulouse och USA där han jobbade för Motorola och Freescale i Phoenix och Chicago. Hans marknadsföringskompetens sträcker sig från halvledare för trådlöst och konsument till trådlös laddning och industriautomationssystem. Han har en magisterexamen i elektroteknik från Grenoble INPG i Frankrike och en magisterexamen i industriella system från San Jose State University.

Bruce Tannenbaum leder Mathworks tekniska marknadsföring inom bildbehandling och tillämpningar för datorseende. Tidigare i sin karriär var han produktchef på halvledarföretag med anknytning till bildanalys, som SoundVision och Pixel Magic, och utvecklade algoritmer för datorseende och wavelet-baserad bildkomprimering på SRI (Sarnoff Corporation). Han har en BSEE-examen från Penn State University och en MSEE-examen från University of Michigan.

För att robotar ska kunna interagera på ett meningsfullt sätt med objekt i sin omgivning och kunna röra sig själv, behöver de kunna se och tolka vad de ser omkring sig. Drömmen om autonoma, adaptiva robotar är gammal. Idag blir den verklighet med hjälp av avancerade – och kostnadseffektiva – bildprocessorer som matas från 3D-sensorer och kör robusta algoritmer.

Robotar, som de länge framställts i filmens värld (och reklamens), lovar att befria människor från monotona, tråkiga och på andra sätt icke önskvärda arbetsuppgifter. Och på köpet höja kvaliteten på utfört arbetet med sin snabbhet och millimeterprecision.

Den första vågen autonoma konsumentrobotar är ett bra exempel. De kan dammsuga, tvätta mattor och till och med rensa hängrännor från höstlöv. Ett annat exempel är de robotar som vi ser vid allt fler tillverkningslinor i olika sorters fabriker.

DEN FÖRSTA GENERATIONEN autonoma konsumentrobotar är dock ganska primitiva varelser när det gäller hur de bär sig åt för att observera och navigera sin omgivning.

Ett av knepen är att sätta upp virtuella hinder i form av IR-sändare som robotens IR-sensorer kan koordinera sig mot – så att den inte ramlar nerför trappan eller går in i fel rum.

Ett annat knep är att bygga in en stötsensor så roboten märker att den har kolliderat med ett stillastående objekt och därmed inte bör fortsätta åt samma håll, eller – om den är mer avancerad – att notera på en inbyggd karta att den inte ska komma tillbaka till samma plats igen.

En industrirobot är visserligen överlägsen människan vid det löpande bandet när det gäller precision, tempo och ork. Men den är helt beroende av att de arbetsstycken som glider fram är rätt roterade och exakt placerade. Detta gör produktionstekniken komplex och minsta avvikelse ger leder till fel i monteringen.

Vi människor använder ögonen (i huvudsak) och hjärnan för att se och navigera. Robotar kan i princip göra detsamma med hjälp av kameror, processorer och smarta algoritmer.

Historiskt har också den typen av bildanalys förekommit, men bara i ett fåtal komplexa dyra system. Idag har priset på kretsar sjunkit. Likaså deras strömförbrukning. Samtidigt har prestanda gått upp.

Därmed har möjligheten till avancerad bildanalys öppnats inom flera högvolymstillämningar, och det har börjat dyka upp allt fler seende robotar.

Visst – det finns en del utmaningar kvar för dem som implementerar systemen. Men aldrig förr har det funnits så enkla, snabba och kostnadseffektiva verktyg för att ta sig an dessa utmaningar. [1]

Robotiserat seende kräver algoritmer som konverterar data från bildsensorer till information som går att omsätta i handling.

En vanlig typ av robotuppgift är att identifiera externa objekt och deras orientering. En annan är att bestämma sin egen position och orientering.

Många robotar är konstruerade för att interagera med en eller flera återkommande identiska objekt. En robot som kan detektera dessa objekt trots att deras position och orientering varierar, och trots att de kanske till och med rör sig, kallar vi för en adaptiv robot.

KAMERAR KAN PRODUCERA miljontals bildpunkter per sekund. Det ger roboten en hög arbetsbörda. Ett vanligt sätt att minska bördan är att först identifiera karaktäristiska bildelement eller kännetecken i bilden, som hörn, homogena regioner (blobs), kanter och linjer.

En sådan transformation – från punkter till bildelement – minskar beräkningsbördan med en faktor tusen eller mer. Miljoner punkter reduceras till några hundra karaktäristiska bildelement som roboten sedan pusslar samman till kompletta objekt med position och orientering.

Autonoma kundanpassade produkter och industriella produktionssystem är några av de många klasser av robotar som kan förbättra sin funktion via bildanalys.

Första steget i att identifiera objekt är att kombinera grupper av bildelement med hjälp av maskininlärning eller andra algoritmer. Efter att man jämfört kombinationerna med en databas av objekt fotograferade ur olika synvinklar och roterade på olika sätt, kör roboten en klassificeringsalgoritm och tränas i att korrekt identifiera nya objekt.

En av de mest kända objektigenkänningsagoritmerna heter Viola-Jones framework. Den arbetar med bildelement av en typ som kallas Haar, och utnyttjar en klassificeringsteknik som kallas Adaboost. Viola-Jones är särskilt bra på att känna igen ansikten men kan också tränas att känna igen andra typer av objekt.

En nackdel med metoder baserade på maskininlärning är att att de behöver stora volymer träningsdata innan deras klassificeringar börjar bli korrekta.

För att bestämma hur ett objekt är orienterat kan man använda algoritmer som RANSAC (Random Sampling and Consensus) som baseras på statistik. Några typer av bildelement väljs ut och används för att modellera orientering, varefter algoritmen undersöker hur många av de övriga bildelementen som matchar modellen.

Den modell som matchar flest bestämmer vilken objektrotation som är korrekt.

FÖR ATT KLASSIFICERA OBJEKT som rör sig behöver man komplettera med en spårningsalgoritm. Efter objektidentifieringen applicerar man algoritmer som KLT (Kanade-Lucas-Tomasi) eller Kalmanfiltrering för att spåra bildelement mellan bildrutor.

Algoritmerna fungerar även när objekt byter orientering eller tillfälligt döljs eftersom man endast behöver spåra en delmängd av elementen.

Algoritmerna ovan kan vara tillräckliga för en stationär robot.

När det gäller rörliga robotar måste man addera ytterligare algoritmer. En kategori av sådan algoritmer heter SLAM (Simultaneous Localization and Mapping). SLAM bygger kartor av omkringliggande miljö samtidigt som den håller reda på robotens egen position. För att det ska fungera krävs att kartläggningen är tredimensionell.

Flera djupdetekterande sensorbaserade metoder finns att välja mellan.

En är att härma människans ögon – det vill säga att konfigurera två enkla kameror till en stereokamera. Sådana använder så kallad epipolär geometri – 3D-koordinater för punkterna i scenen härleds genom projektioner från de två 2D-bilderna.

Bildelement kan inte bara användas för att analysera 2D-bilder utan också för att för att att detektera intressanta objekt i en 3D-scen.

Exempelvis är det mycket enklare för en robot att detektera kanten av ett bord än en plan väggyta.

ALLTEFTERSOM ROBOTEN rör sig eller roterar, fortsätter den att detektera bildelement och jämföra med och uppdatera den interna karta den bygger upp för att lokalisera sig själv. Med tanke på att objekt i verkligheten ofta förflyttar sig, är en statisk karta sällan till nytta för en robot som vill anpassa sig till sin omgivning.

När vi diskuterar effektiva implementeringar av robotseende, är det lämpligt att först dela upp ovannämnda analyssteg i fyra faser.

Varje fas har egna unika kännetecken och bivillkor vad gäller den beräkningskraft som krävs. [1]

Flera olika typer av processorer används för bildanalys. De är olika lämpade för de olika algoritmfaserna vad gäller bland annat prestanda, energikonsumtion, pris, och flexibilitet.

En och samma bildprocessor kan innehålla flera olika sorters beräkningskärnor för att adressera olika unika behov i beräkningsfaserna.

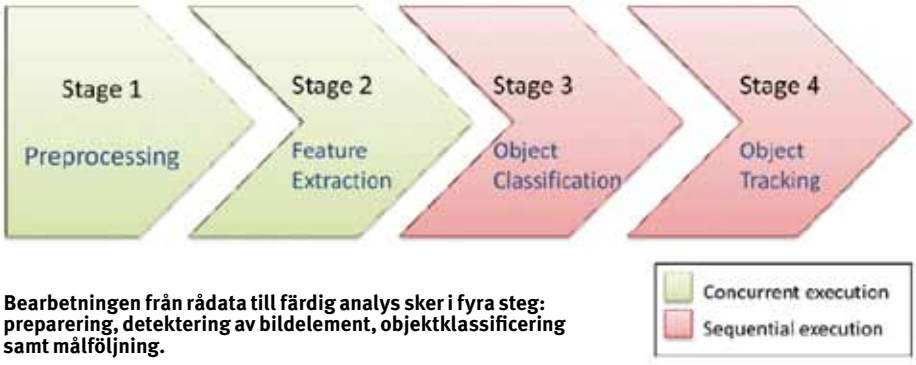
I den första fasen hittar vi algoritmer som preparerar sensordata på olika sätt:

- Skalning
- Färgrumskonvertering
- Bildrotering och invertering
- De-interlacing
- Färgjustering och mappning mellan färgomfång
- Gammakorrigering, och
- Kontrastering

KÄNNETECKNANDE ÄR ATT varje enskild bildpunkt bearbetas vilket resulterar i ett enormt antal operationer per sekund. I fallet med stereovision processas båda bilderna parallellt.

Ett av alternativen är att använda ett dedikerat hårdvarublock, ofta kallat kallat IPU (Image Processing Unit). De senaste bildprocessorernas IPU:er kan hantera dubbla bilder med en upplösning på 2048×1536 punkter (drygt tre miljoner) i en stadig bildfrekvens.

Den andra fasen detekterar tidigare nämnda bildelement, eller kännetecken, ▶



Bearbetningen från rådata till färdig analys sker i fyra steg: preparering, detektering av bildelement, objektklassificering samt målföljning.

En processor för bildanalys kan innehålla olika sorters beräkningskärnor för att hantera de olika beräkningsstegen.

genom att extrahera hörn, kanter och andra intressanta bildregioner.

Också detta sker på basis av en bildpunkt i taget, vilket betyder att man föredrar massivt parallella arkitekturer som i detta fall måste kunna hantera lite mer komplexa matematiska funktioner, som första- och andraderivata.

Både DSP:er, FPGA:er, GPU:er (grafikprocessorer), IPU:er och APU:er (array processor units) är möjliga processoralternativ.

DSP:er och FPGA:er är flexibla och därför lämpliga under den tid när tillämpningar och algoritmer fortfarande är omogna och utvecklas.

Däremot är de mindre konkurrenskraftiga vad gäller energiförbrukning, prestanda och pris.

I ANDRA ÄNDEN av spektrumet mellan specialisering och flexibilitet, finns IPU:er och APU:er med inbyggda operationer för bildanalys.

De kan utföra miljardtals operationer per sekund, men är hårt optimerade för vissa tillämpningar och ger inte bred funktionalitet som DSP:er och FPGA:er.

I mitten av nämnda spektrum hittar vi grafikprocessorn, GPU:n. Den användes historiskt i datorer men finns idag också i approcessorer i smarttelefoner, pekplattor och andra högvolymsprodukter.

GPU:n är väl lämpad för flyttalsberäkningar, som exempelvis minsta kvadratmetoden i optical flow-algoritmer, i deskriptorberäkningar i SURF-algoritmen (Speeded Up Robust Features algorithm, som används för snabb igenkänning av nyckelpunkter) och i punktmolnsberäkningar.

Ett alternativ till GPU är SIMD-vektorprocessormotorer som ARM:s NEON eller Altivec-funktionsblocket i Power-CPU:er.

För nämnda hårdvaror hittar man lämpliga mjukvaruverktyg i ramverk och bibliotek som OpenCL (Open Computing Language) och OpenCV (Open Source Computer Vision Library). Dessa förenklar och snabbar upp kodutvecklingen. De tillåter också att man delar upp och allokerar beräkningsuppgifter på olika kärnor.

I DEN TREDJE FASEN ska systemet klassificera objekt baserat på mönster av bildelement. Här är bearbetningen inte längre punktbaserad. Algoritmerna är istället mycket icke-linjära i sin struktur och i sin access av data.

De kräver dessutom mycket beräkningskraft om det är många olika kännetecken som ska matchas mot en stor databas.

Här är konventionella risc-processor det ideala valet – exempelvis Power eller ARM-CPU:er i enkel- eller multikärnor.

Detsamma gäller fas nummer fyra, där man spårar klassificerade objekt mellan bildrutor, konstruerar en modell av miljön och bestämmer huruvida det föreligger några situationer som kräver åtgärder.

Generellt i alla faser gäller med tanke på hur dataintensiv bildanalys är att du när du väljer processor inte bara ska titta på antalet kärnor och prestanda per kärna, utan också vilken kapacitet att skyffla data den har, exempelvis i termer av extern minnesbussbandbredd.

Allteftersom kapaciteten ökat i processorer, bildsensorer, minnen och andra halvledare, och algoritmerna mognat har man kunnat börja inkorporera avancerad datorbildanalys i allt fler typer av inbyggda system.

Med ”inbyggt system” menas här mikroprocessorbaserade system utom generella datorer. Men ”inbyggt datorseende” avser vi implementationer av datorbaserad bildanalys i inbyggda system, mobila enheter, specialbyggda PC och i molnet.

Datorseende i inbyggda system (embedded vision) kan potentiellt användas för att göra många typer av elektroniska produkter (som de robotiserade system som nämns i denna artikel) intelligentare och mer responsiva än tidigare, och därmed värdefullare för sina användare. Tekniken kan addera värdefulla funktioner till befintliga produkter.

Dessutom kan tekniken skapa nya marknader för dem som tillverkar hårdvara, mjukvara och halvledare.

Embedded Vision Alliance – en världsomfattande organisation bestående av teknikutvecklare och -leverantörer – vill göra det möjligt för ingenjörer att förvandla dessa potentialer till verklighet.

Freescale och Mathworks som tillsammans skrivit denna artikel, är medlemmar i Embedded Vision Alliance vars primära uppdrag är att tillhandahålla utbildning, information och insikter till ingenjörer för att hjälpa dem att inkorporera datorseende i nya och existerande produkter.

FÖR ÄNDAMÅLET har organisationen utvecklat en webbplats (www.Embedded-Vision.com) där man kan hitta lektioner, video och källkod för nedladdning, och ett diskussionsforum med tillgång till vitt skilda tekniska experter.

Det kan också vara en god idé att besöka Alliansens endagsforum Embedded Vision Summit som nästa gång hålls i maj i Santa Clara i Kalifornien. Målgruppen är ingenjörer som är intresserade av att inkorporera visuell intelligens i elektroniska system och till mjukvara.

På agendan står how-to-presentationer, seminarier, demonstrationer och möjligheten finns att interagera direkt med Alliansens medlemsföretag.

ATT OMVANDLA EN IDÉ om robotics vision till en skeppad produkt kräver både gott omdöme och kompromissvilja.

Embedded Vision Alliance är en katalysator för samtal och ett forum där man möter snabb förståelse för avvägningar och hur man gör dem.

Alliansen hjälper till att accelerera satsningar på produktifiering av avancerade robotsystem och tillåter systemutvecklare att effektivt ta i bruk teknik för datoriserad bildanalys. ■

[1] “Embedded Low Power Vision Computing Platform for Automotive” Michael Staudenmaier, Holger Gryska, Freescale Halbleiter Gmbh, Embedded World Nuremberg Conference, 2013.