
Miljöanpassning av gestidentifieringsteknik i bildanalys

Måns Jarlskog

Att med hjälp av en dator, automatiskt och snabbt, kunna känna igen och lokalisera ansikten, handgester och andra objekt i digitala bilder är en teknologi som blir allt viktigare.

Ovannämnda arbetsuppgifter är lätta för enskilda människor att utföra. Genom att ge datorn samma förmåga kan stora mängder bild och videomaterial analyseras och tekniken är t.ex. helt avgörande för framtidens självkörande bilar. Gestigenkänningstekniken har också gjort det möjligt för människor att interagera med mobiler och surfplattor endast med hjälp av den inbyggda kameran. Den underliggande tekniken kan också användas för mer avancerade uppgifter t.ex att, från ultraljudsbilder, tidigt hjälpa till att upptäcka defekter i hjärtat och andra organ men också finna mönster i stora datamängder (data-mining).

Maskininlärning

Gemensamt för tekniken är att den är baserad på maskininlärning som går ut på att ge en dator förmågan att lära sig att ta önskade beslut utan att vara helt förprogrammerad. En vanlig träningsmetod är att leta efter och samla på mönster som sedan tillsammans används för att rösta fram tillräckligt korrekta resultat på data med facit, s.k. träningsdata. I detta tillvägagångssätt är valet av mönster-typen och träningsdatan helt avgörande.

Problem

Ett vanligt problem är att något användningsområde bland de avsedda användningsområdena är underrepresenterad i träningsdatan. Detta kan i sin tur leda till att detektorerna underpresterar i den under-representerade miljön. Sådana problem upptäcks dessvärre oftast, under användning, efter träningen. Vanligtvis åtgärdas problemet då genom att samla in mer data och utföra en omträning. Att träna en detektor kräver mycket datorkraft och tid och därför är det attraktivt att istället kunna anpassa den befintliga detektorn.

Miljöanpassning

I examensarbetet har jag utvecklat en metod som låter förtränade handgest-detektorer själva upptäcka svåra miljöer under tiden de används och därefter erbjuda användaren

en möjlighet att anpassa detektorn till den svåra miljön. Dessa gest-detektorer är avsedda att leta igenom flera bilder i sekunden från t.ex. mobilkameror.

Mitt examensarbete består av tre delar. Först inför jag en algoritm som tillåter detektorn att själv upptäcka en väg (men) representerad miljö. Sedan, en algoritm som uppskattar vilka, av de befintliga, mönster som är relevanta i den svåra miljön. Sista delen består av algoritmer som föreslår nya miljöanpassade röstvikter till de befintliga mönsterna. För att sedan adaptera detektorn används de anpassade röstvikterna istället för de gamla. Det är mycket möjligt att de nya röstvikterna inte ger tillfredsställande resultat utanför miljön de är anpassade för. Därför är de anpassade röstvikterna konstruerade så att det snabbt går att ta bort och spara dem. Att man kan spara undan vikterna ger en möjlighet att återanvända en tidigare lyckad adaptation.

Resultat

Jag testade metoden ovan på fyra förtränade detektorer, avsedda för att hitta öppna händer, som underpresterar i miljöer med svagt ljus. Detektorerna samlade data från videosnuttar, från en miljö med svagt och varierande ljus, motsvarande 4 minuters exponering. För att inte förstöra den befintliga detektorn, användes de nya röstvikterna endast till redan klassade men osäkra exemplar. Detta visade sig dock överflödigt eftersom de nya röstvikterna ändå inte tycks påverka de exemplar som klassades med större säkerhet. Resultatet blev fler upptäckta händer med bara några få procent tyngre beräkningar, men även någon enstaka falsk detektion. För att dra ytterligare slutsatser om hur lyckad adapteringen är och hur den påverkar prestandan i övrigt behövs mer data och lämpliga test-verktyg.