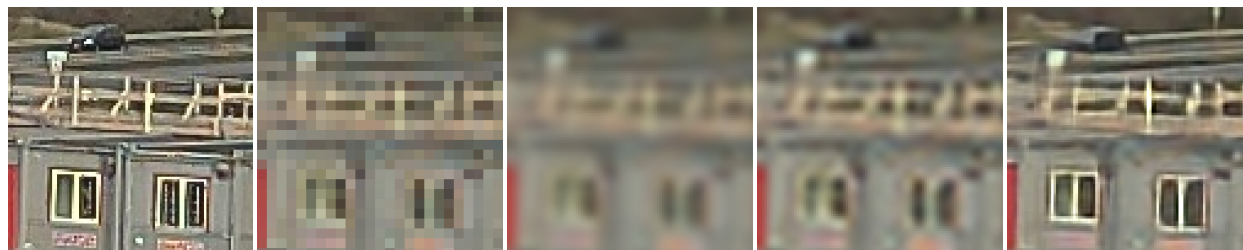


Kan man lära en videokamera att minnas vad den sett?

Johan Förberg, Lunds tekniska högskola

6 december 2017



Originalbild

4 ggrs. nedskalning

Konventionella metoder

Vår metod

Många kameror sitter fast monterade och filmar samma plats varje dag. Tänk om kameror kunde tränas till att minnas vad de sett?

Trots att Sveriges fibernät ständigt byggs ut räcker hastigheten sällan till. Om vi tittar på HDTV idag så vill vi titta på 4K-TV imorgon och detta ställer allt högre krav på vår internetförbindelse. Ännu värre är det för den som har många internetuppkopplade videokameror—det blir som att ha dussintals HDTV-apparater som alltid är igång. Istället för dyra utbyggnader av infrastrukturen jobbar vissa ingenjörer med att minska mängden data som krävs för att beskriva en videoström.

Den som hittar en bild på en webbsida och försöker skriva ut i stort format blir i regel besviken på det otydliga och "pixliga" resultatet. Den nedladdade bilden är för lågupplöst—fotografen har minskat på upplösningen för att bilden ska ta upp mindre plats. Likadant gör man med video och resultatet blir detsamma: pixligt eller suddigt. Varför då?

En digital bild består av många pixlar som var och en beskriver vilken nyans bilden har i en viss punkt. När man minskar upplösningen kastar man bort vissa pixlar, exempelvis kan man halvera upplösningen genom att endast spara varannan pixel. De pixlar som kastats bort måste oftast återskapas på något sätt innan bilden kan visas. Detta kan man göra genom att räkna ut ett medelvärde mellan närliggande pixlar vilket motsvarar att färgen smetas ut och gör bilden suddig. Lite smartare metoder bygger på att bilden samtidigt "skärps till" så att kanter och detaljer i bilden framhävs. Detta motverkar delvis suddigheten men kan leda till "ringningar" och andra onaturliga inslag i bilden.

De senaste åren har en helt ny teknik lanserats för att återskapa förlorade pixlar, en teknik som använder så kallad *maskininlärning*. Till skillnad från vanlig datorprogrammering, där en ingenjör beskriver för datorn exakt vad som skall göras steg för steg, bygger maskininlärning på att datorn själv skall formulera en lösning. Ett sådant system innehåller en uppsättning bildfilter som kan konfigureras genom tusentals olika parametrar. Genom att ändra på parametrarna kan filtren ställas om till att göra många olika saker. Under en flera timmar lång träningsession låter man systemet se tusentals före- och efterbilder för den operation som man vill träna in. I varje steg justerar systemet sina filterparametrar en aning så att man med tiden kommer närmare och närmare den önskade operationen.

Med tidigare metoder har man försökt maximera mångfalden bland träningsexemplen i förhoppning om att systemet efter träning ska kunna hantera många olika sorters bilder. Med vår metod låter man istället en fast monterad kamera träna på bilder av samma plats vid olika tidpunkter. Med tiden kan systemet till synes "memorera" de föremål som kameran ser vilket gör det möjligt att återskapa dessa föremål med högre noggrannhet än tidigare. På så vis sparar man på sin internetuppkoppling då det krävs färre pixlar för att beskriva bilden.

Den som inte är bekymrad över datamängd kan använda systemet för att återskapa pixlar som aldrig har funnits. Då kan man få en bild som har högre upplösning än vad kameran egentligen klarar av. Resultatet blir något som påminner om amerikanska deckarserier eller Harrison Ford i *Blade Runner*, även om det i verkligheten finns gränser för vad som går att återskapa i efterhand.