

Az utóbbi évtizedekben tapasztalható drasztikus teljesítmény-növekedésnek köszönhetően számtalan feladat megoldását tudjuk számítógépekkel végezni: folyadékok viselkedését szimulálhatjuk, élethű virtuális fényképeket tudunk készíteni, vagy éppen tizedmásodpercek alatt hozzájuthatunk tetszőleges információhoz egy világméretű hálózaton. Ha a számítógépek ilyen sokféle és nagyméretű adathalmazt képesek feldolgozni (azaz megérteni), jogosan merül fel a kérdés: képesek-e látni? Meg tudja-e például mondani egy számítógép, hogy mit ábrázol a betöltött kép vagy videó? Tudunk-e a közeljövőben „keresőszavas” keresés helyett „keresőképes” keresést indítani? Milyen jó is lenne, ha a családi videók között gombnyomásra megkaphatnánk azokat a felvételeket, melyeken Józsi bácsi Márta néniével, de Zolika nélkül szerepel!

A gépi látás (computer vagy machine vision) tudományterület pont ezekre a problémákra keres megoldásokat. Bár arra egyelőre még esélyünk sincs, hogy számítógépeink úgy „általában” megértsék, mit ábrázolnak a képek, adott kérdésekre sikeresen kaphatunk választ. A különféle módszerek speciális szempontok alapján próbálnak meg információt kinyerni az álló- és mozgóképekből (sőt, akár 3 dimenziós képekből is). A gépi látás témaköre igen tág és sok ponton kapcsolódik egyéb tudományterületekkel. A teljesség igénye nélkül tekintsük át a legfontosabb feladatokat és témaköröket:

Orvosi képfeldolgozás: A mikroszkópos, ultrahang, röntgen, mri, ct stb. leképezések számítógépes analízisének során megkísérljük (többek között) az abnormális elváltozásokat (pl. tumor, daganat) automatikusan felismerni, valamint a szövetek közötti határvonalak (vagy felületek) azonosításával a képet szegmentálni. Egyelőre nem kell attól tartanunk, hogy emberi felügyelet nélkül, számítógépek utasítására kerülünk kés alá, hiszen ezek az adatok csupán segítik a szakorvosok munkáját. Bár az algoritmusok nem tökéletesek, szerencsére viszonylag pontosan megmondható a tévedés valószínűsége. Az automatikus analízis eredménye soha nem határozott pozitív vagy negatív diagnózis, hanem ezek százalékos valószínűsége. A szegmentációs eljárások jól illusztrálják, hogy a képek feldolgozása nemcsak 2 dimenzióban történhet. Bár a röntgenfelvételeken is nagyon hasznos a csontok

Gépi látás

I. rész



határának pontos felismerése, a gyakorlott orvosnak erre nincsen feltétlenül szüksége. Sajnos 3 dimenziós adatoknál már nem ez a helyzet: bár a térbeli adatok megjeleníthetők „szeletek” formájában, ez sokszor nem megfelelő. A számítógépes feldolgozás során előállíthatók – a 3 dimenziós sűrűségi adatok szegmentálásával – a szöveteket határoló felületek, így lehetőség nyílik a szervek megtekintésére a virtuális 3D térben.

Ipari minőségellenőrzés: Kamerák segítségével hatékonyan kiszűrhetők a hibás termékek szín, forma vagy egyéb látható jellemző alapján. Ez az emberek számára fárasztó és monoton munka különösen hatékonyan automatizálható. A biológiai látásnál bizonyos szempontból még pontosabb is a „gépi látás”, hiszen az akkomodáció (környezethez való alkalmazkodás) miatt torzulhat érzékelésünk. Gondoljunk csak egy egyszerű gyakorlati példára: számtalan minőségi előírásnak kell megfelelnie egy paradicsomnak ahhoz, hogy „extra” minőségi osztályba tartozzon. Ezen jellemzők mérése (szín, forma, felületi sérülések) teljesen automatikusan végezhető számítógép segítségével.

Katonai alkalmazások: A gépi látás területén végzett kutatások eredményeinek legnagyobb hasznosítója valószínűleg az amerikai hadsereg, de természetesen erről

nagyon kevés információ kerül a nyilvánosság elé. Épületek, harci és polgári járművek, katonák és egyéb objektumok felismerése több területen is alapvető fontosságú: légi és műholdas felvételek „offline” elemzésénél, és a vadászpilóták pilótáinak harc közben történő segítségénél egyaránt. Akárcsak az orvosi alkalmazásoknál, itt is jól körülírható, hogy mit keresünk a képen. A repülőterek és rakétasilók detektálására optimalizált algoritmus természetesen teljesen más módon működik, más követelményeknek felel meg, mint a repülőgépek fedélzetén használt „automatikus célpont azonosítók”. Természetesen ezen a területen is a hagyományos fényképek és videojelek mellett igen sokféle adatot használnak az elemzések során, többek között radar, infravörös kamera, lézeres letapogató vagy éppen vízalatti szonár által szolgáltatott jeleket.

Robotok irányítása: A robotok önálló navigációja már nagyon régóta kutatott terület, az eredmények napjainkban jelennek meg a legkülönbözőbb termékekben. Az alapvető feladatok ezen a területen: repülő/guruló/úszó járművek irányítása, akadályok elkerülése, vészhelyzet felismerése. Bár valószínűleg még sokáig csak science-fiction marad az autókban használható autópilóta, már több forgalomban lévő autó rendelkezik olyan rendszerrel, mely figyelmeztet a

sáv elhagyására. Szintén felhasználják a gépi látás eredményeit, hogy a vezetőt figyelmeztessék az esetleges ütközés veszélyére, netán be is avatkozik az elektronika a vészhelyzet elkerülése érdekében.

Kamerás megfigyelés: Gyakorlati alkalmazások szempontjából szintén kiemelten fontos területek az objektum felismerés és követés. Nem csak rendőrségi és biztonsági alkalmazásai vannak a járókelők, arcok, mozgó járművek detektálásának és követésének, hiszen pl. napjaink egyik „legforróbb” kutatási területe a forgalomszámlálás. A forgalomról szerzett információk birtokában optimálisabban vezérelhetők a közlekedési lámpák, és az autók is ügyesebben megtervezhetik útjukat a városban. Gyalogosok számlálásának is vannak gyakorlati alkalmazásai, hiszen így például lehetővé válik annak pontos mérése, hogy hány járókelő néz meg egy adott plakátot.

Interakció számítógépekkel: Bár sokan sokféle eredményeket értek el a számítógép előtt ülő személy tekintének, arckifejezésének vagy kézmozdulatának követése kapcsán, a gyakorlati elterjedés – úgy tűnik – még várat magára. Valószínűleg még egy ideig egyszerűbb lesz az egérrel „odabököni” valahova, mint megpróbálni úgy ránézni egy ikonra és úgy kimondani az utasítást, hogy azt a számítógép bizonyosan megértse. Érdekes módon a kézmozdulatok felismerésének egyik legelső alkalmazására a televíziós időjárásjelentéseknél lelhetünk: bizonyos csatornáknál már speciális információkat jelenítenek meg a képen attól függően, hogy a térképen pontosan hová mutat meteorológus.

Karakterfelismerés: Talán nem kell hangsúlyoznunk a téma jelentőségét. A nyomtatott és a kézzel írott írás felismerése is régóta kutatott téma, igen jelentős, és a gyakorlatban is bevált eredményekkel. Gondoljunk csak a postai küldemények címezésének felismerésére, kitöltött nyomtatványok automatikus feldolgozására, vagy könyvek és dokumentumok gyors digitalizálására.

Sztruktúra és mozgás rekonstrukciója: Talán a legkomplexebb feladat a gépi látás világában az úgynevezett „structure from motion” problémájának megoldása. A feladat: képsorozat vagy videó alapján az eredeti kamera mozgásának, és a felvételen látható objektumok térbeli elhelyezkedésének kiszámítása. Gondoljunk csak a modern digitális trükkökre: eredeti, azaz élő felvételekbe helyeznek el számítógépes grafikával létrehozott épületeket, járműveket, embereket, óriásmajmokat vagy éppen orkok hadseregét. Ezt csak úgy lehet kivitelezni, ha a virtuális szereplőket pontosan olyan virtuális

kamerával fényképezzük, mint az eredeti jelenetet, így a két anyag kombinációja illeszkedni fog. Ezen algoritmusok „merek” környezetben működnek hatékonyan, tehát feltételezzük, hogy a kamera által látott környezet nem változik. Hasonló feladat az élő szereplők mozgásának rögzítése, azaz a motion capture is. A feladat itt annyiban más, hogy ismert (kalibrált) kamerákkal rögzítjük a mozgó szereplőket és tárgyakat, majd ezek mozgását próbáljuk meg rekonstruálni. Itt azonban már messze nem merek a környezet, így több kamerára is szükség van. A szereplők mozgásának kitalálásánál figyelembe kell venni rengeteg megkötést, hiszen tudjuk például, hogy a csontok nem nyúlnak és nem rövidülnek meg mozgás közben.

Képek szegmentálása: A gépi látás szinte összes alkalmazásánál az egyik első lépés a szegmentálás. Ebben a lépésben valamilyen szempont alapján izoláljuk a kép egy (vagy több) részét a többitől, leggyakrabban a fontos és a felesleges területeket választva szét. A szegmentálást számtalan jellemző kombinációja alapján végezhetjük: szín, világosság, textúra, mozgás stb. Gyakran nem magát a kiválasztani kívánt területet keresik az eljárások, hanem a területek közötti határvonalakat (éldetektálás).

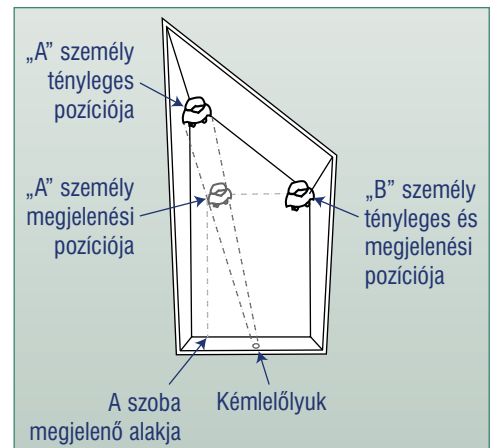
Keresés képeket tartalmazó adatbázisban: bár a számítógépek még nem alkalmassak a képek felismerésére úgy általában, kellően behatárolt feladatkörben jól alkalmazhatóak. Az emberi arc vagy ujjlenyomat felismerése jó példa erre: bizonyos – előre megadott – jegyek alapján könnyen megállapítható algoritmikusan, hogy két arc/ujjlenyomat hasonlít-e vagy sem. A felismerés alapjául szolgáló jellemzők (feature) kinyerése kulcsfontosságú, ha ez jól történik meg, nem nehéz a felismerés.

Sajnos az emberi látás, pontosabban a látott képek értelmezése rettenetesen összetett feladat. A legtöbb ember azt gondolja, hogy agyunk – akár egy számítógép – analizálja a szemünk által érzékelt képet. Kiszámolja a tárgyak távolságát, megmondja, hogy mi piros és mi fehér, vagy egy arc elemzésével megmondja, hogy ismerős-e az illető vagy sem. A dolog sajnos (vagy szerencsére?) nem egészen így működik. A látás alapvetően tanult folyamat, és a kép(részletek) felismerésén alapszik. Elég a mellékelt „Ames szoba” képre néznünk, rögtön látszik, hogy agyunk tévesen azonosítja a látott helyiség dimenzióit! Természetesen nem a kislány nőtt irdatlanul nagyra, csupán a szoba alakja egyedi!

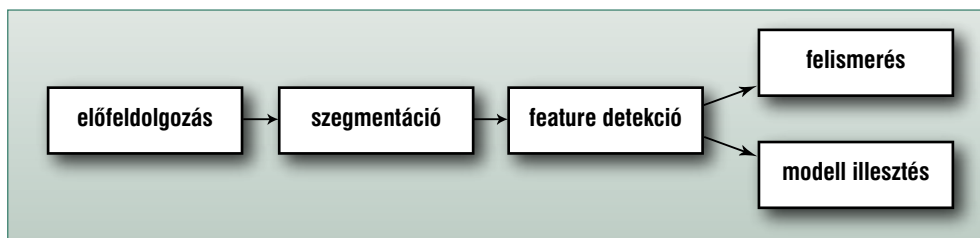
Ha agyunk csupán matematikai egyenletek megoldásával próbálná értelmezni a képeket, ...hát nem sokra jutna. Egyik szemünk becsukásával, elvileg képtelenek lennénk megmondani az előttünk lévő tárgyak távolságát. Mivel azonban agyunk „megtanulta”, hogy a közelebbi tárgyak nagyobbak, mint a távoliak, könnyedén tájékozódunk fél szemmel is.

A következőkben áttekintjük, hogy a fenti feladatokat a számítógép segítségével hogyan lehet megoldani. Bár elsőre nem szembetűnő, de ezen problémák megoldása hasonló lépésekkel történik.

Szinte mindig célszerű a képeket egyszerű transzformációkkal, szűrőkkel előfeldolgozni. Az így kapott kép valószínűleg tartalmaz fontos és lényegtelen területeket is. Ezeket a szegmentáció során választjuk szét egymástól. Így megkaptuk a kép számunkra fontos területét, de ez még mindig pixelek ezreit (akár millióit) jelentheti, ami a legtöbb algoritmus számára túl sok feldolgozandó adatot jelent. A sűrű képpontfelhőből kezelhető mennyiségű adatot az ún. feature-ök, vagy „vizuális jellemzők” kinyerésével kapunk. Ez legtöbbször vonalak, sarkok vagy egyéb szempontból speciális elemek detektálását jelenti. Így már nem pixelekkel, hanem kezelhető mennyiségű képi jellemzővel kell dolgoznunk. Ezek alapján már megkereshet-



Ames szoba



Az egymást követő lépések a gépi látás megvalósításában

jük az előre eltárolt adatbázisban a jellemzőkhöz legjobban hasonlító objektumot (felismerés), vagy kiszámíthatjuk a legjobban illeszkedő (mozgás, pozíció) paramétereiket. Ne felejtjük el azt sem, hogy a képi jellemzők kinyerése során hamis találatokat is kaphatunk, amiket fel kell ismerni, és ki kell szűrni az adathalmazból. A következőkben, gyakorlati példákon keresztül, ezekkel a lépésekkel fogunk részletesebben megismerni.

Előfeldolgozás

Az előfeldolgozás célja a nyers bemeneti adat előkészítése egyszerű műveletek és szűrők segítségével. Általában a két legfontosabb szempont a főleg az információk eldobása, kiszűrése valamint a bemeneti képeken megjelenő nem lényeges különbségek kiküszöbölése. Ez utóbbira jó példa az automatikus célpontfelismerés: napsütésben (erős kontraszt, világos kép), felhős időben (gyenge kontraszt) és éjszaka is (sötét kép) fel kell ismernünk ugyanazt az objektumot. Éppen ezért a legelső lépés a világosság és a kontraszt azonos szintre állítása, így a későbbiekben ez nem okozhat problémát. Mint azt később látni fogjuk, legtöbbször a képen látható sarkokkal vagy éllel dolgozunk, ezek azonosításához pedig nem feltétlenül szükséges a színinformáció. A főleg az információk eldobása éppen ezért gyakran nem más, mint a bemeneti kép átalakítása szürkeárnyaltos képpé. Ez első látásra egyszerűnek tűnhet, de a valóságban nem árt körültekintőnek lenni. Gyakran éppen a képpontok színe kódolja a képi információ javát. Gondoljunk csak a gépi minőségellenőrzés példájára: ha például barnás forradásokat

keresünk kék szilvákon, érdemes a bemeneti képek B, azaz a kék csatornával dolgozni, a többi pedig eldobni. Így a foltok jól azonosítható, teljesen sötét területek lesznek.

Szegmentálás

Bár az előfeldolgozás az értékes információk kiemelését, és a felesleg eldobását szolgálja, ezt azonban csupán a képpontok

színének és világosság értékének módosításával teszi. A szegmentáció során a kép jellemző területeit választjuk szét, legtöbbször a „felesleg” és „érdekes” halmazba csoportosítva azokat. A szegmentáció is a legtöbbször szín és világosság alapján szűri a képpontokat. Navigációs alkalmazásoknál például nem érdekes az égbolt, forgalomszámlálásnál az útfelület és az állandó háttér, műholdas képeknél pedig a felhőkkel takart területek.

A szín alapján történő szegmentációt úgy képzelhetjük el, hogy a 3 dimenziós RGB – vagy éppen az 1 dimenziós szürke-árnyalat – értékek „terében” kijelölünk egy kisebb pontthalmazt, általában kockát vagy paralelepipedont, és a belülről eső pontokat az egyik halmazba, a kívülről esőket a másik

halmazba soroljuk. (A jól ismert kék háttér „kulcsolása” is hasonló módon történik.)

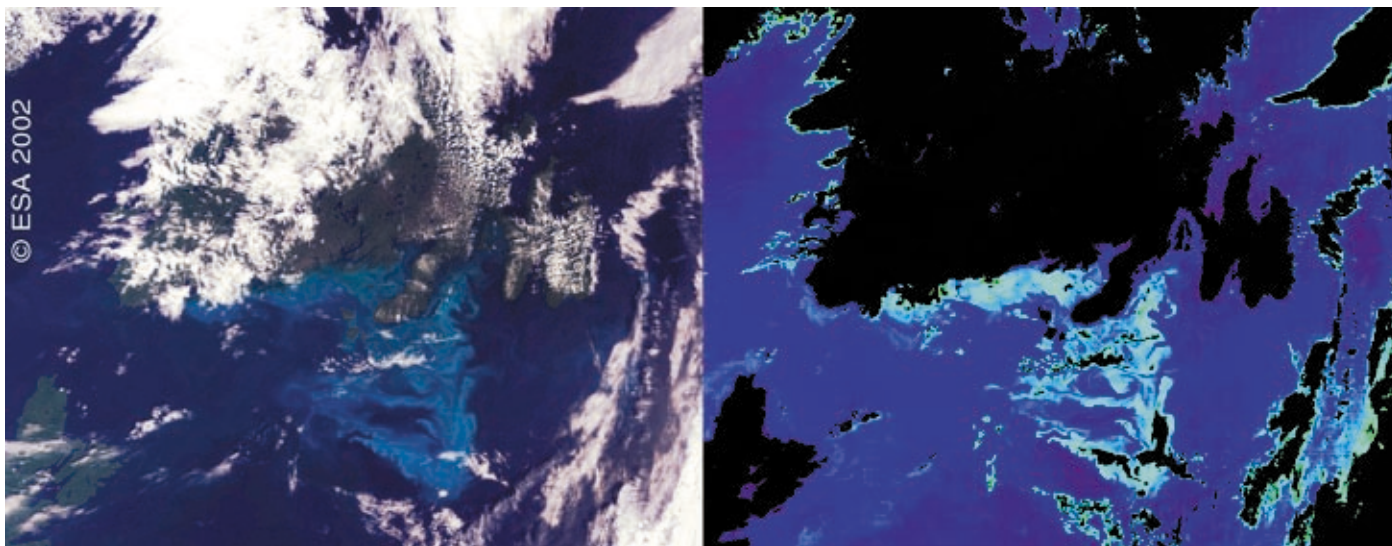
Ha fix kamera által rögzített képeken mozgó objektumokat akarunk detektálni és követni, akkor érdemes a különbség alapú kulcsolást (difference key) alkalmazni. Ennek az a lényege, hogy rögzítünk egy „tiszt” háttérteret, ahol nincsenek járművek, gyalogosok stb., és minden későbbi feldolgozandó képet ezzel hasonlítunk össze. Ahol nincs eltérés, ott nagy valószínűséggel a háttérteret látjuk, ahol jelentős a különbség, ott általában valamilyen objektum van. Így igen könnyedén szegmentálhatjuk az érdekes és érdektelen részeit a képnek. Természetesen rengeteg módszert használnak a gépi látás területén a képek szegmentálására, de ezek ismertetésére még csak kísérletet sem teszünk.

Karakterisztikus jegyek

Az előfeldolgozott és szegmentált képek már közvetlenül alkalmasak lehetnek arra, hogy bizonyos algoritmusok felismerjék vagy kövessék az azokon látható objektumokat. A gyakorlatban azonban ez kevésbé jellemző, ugyanis még így is képpontok ezreit, netán millióit kell feldolgozni, ami a legtöbb módszer számára még mindig kezelhetetlen mennyiség. Szerencsére azonban erre nincs is szükség, hiszen a legtöbb



Eredeti kép, az éldetektált és a szegmentált változat (előfeldolgozás)



Műholdas képeken a szín alapú szegmentáció segítségével kiemelték a tengeri planktonok képét

esetben, a probléma alapos elemzésével jól megfogalmazható, hogy milyen vizuális jellemzők, illetve meghatározó jegyek alapján tudjuk elvégezni a feladatot. Gondoljunk csak arcok felismerésére: butaság lenne közvetlenül a pixelek alapján összehasonlítani két arcot. A hasonlóság mértéke sokkal jobban megfogalmazható különböző biometria értékek – szemek, orr, áll távolsága, arc szélessége, homlok magassága – alapján. Robotok által irányított járművek esetében sem tudunk sok mindent közvetlenül kezdeni a jármű előtt elterülő út – szegmentált, előfeldolgozott – képével. Hogy az úton tudjuk tartani a robot, vagy alvó sofőr által vezetett autót, tudnunk kell, hogy merre van a záróvonal, az útpatka, netán az előttünk haladó jármű. Nyomtatott vagy kézzel írt szöveg, vagy éppen ujjlenyomatok felismerése is a karakterisztikus jegyek, azaz a feature-ök azonosításán és feldolgozásán alapszik. Ezek

a speciális képi jellemzők a legtöbb esetben robusztusan kinyerhetők a képből, azaz még az esetleges hibát, téves detektálást is ki tudjuk védeni. Mint az a példából kiderült, az egyes feladatok jelentősen különböző feature-ök kinyerését igénylik, a leggyakrabban azonban kétségtelenül az élek, kontúrok, sarkok, pontszerű jelek és egyéb, jól azonosítható formák megtalálása a feladat.

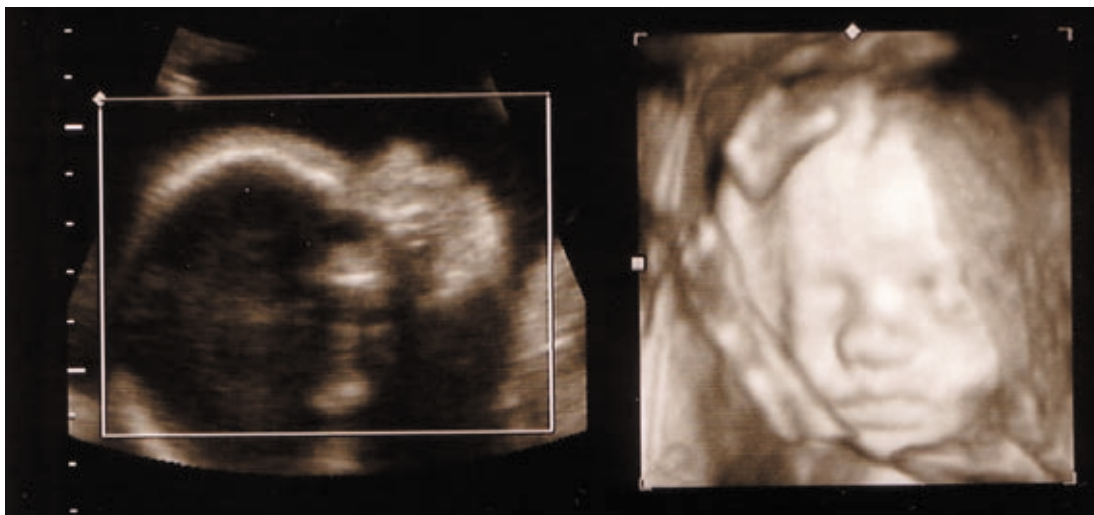
A karakterisztikus jegyek kinyerése képekből már évtizedek óta kutatott terület, így számtalan hatékony, már-már standardnak tekinthető algoritmust ismerünk. A legtöbb módszer a szomszédos képpontok különbségén, azaz a „képfüggvény deriváltján” alapszik. Matematikailag viszonylag jól megfogalmazható, hogy milyen feltételeknek kell teljesülniük az éleken és a sarkokon található képpontokban. Szintén nem nehéz egyenes vonalakat, vonalak metszéspontjait, köröket és egyéb szabályos geometriai for-

mákat megtalálni a képen. Modern alkalmazásoknál – például objektum felismerése és követése – olyan jellemzőket is kinyernek a képekből, melyek nem csupán pozíciójukkal, hanem orientációval vagy mérettel is jellemezhetőek (például detektálják az apró pontokat és a nagyobb „korongokat” is, sőt ezek relatív méretét is). Az sem ritka, hogy egészen pontosan meghatározható dolgokat keresünk a képen: szemeket, kézfejet vagy éppen egy rendszámot. Ilyen esetben akár egészen speciális algoritmusok is rendelkezésre állnak a hatékony feature detekcióra.

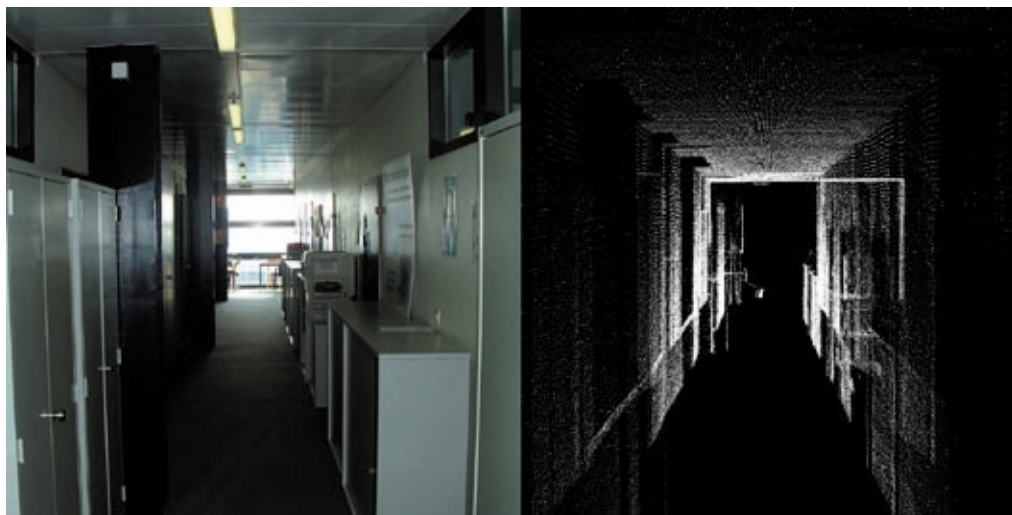
Felismerés és modellillesztés

A karakterisztikus jegyek azonosítása után általában nem több mint néhány száz leíró, avagy descriptor tömöríti magába a feldolgozandó képi információt. Ez már sokkal kezelhetőbb, mint sok millió RGB képpont! Nem meglepő módon a gépi látás különböző alkalmazásai meglehetősen más és más módon dolgozzák fel ezt az információmennyiséget, de a két legnagyobb problémakör kétségtelenül a felismerés és a modellillesztés.

A felismerés – legyen az célpont, arc, autótípus, daganat – a legtöbb esetben nem egy adott objektum felismerését jelenti, hanem egy – akár igen csak terjedelmes – adatbázisban, a legjobban illeszkedő elem megtalálá-



Ultrahangos képek alapján felületek rekonstrukciója



Fotók alapján a felületek rekonstrukciója modellillesztés segítségével

sát. Az arcfelismerés esetében ez azt jelenti, hogy a megtalált leírók – ebben az esetben biometria értékek – sorozatát, azaz feature vector-t kell összehasonlítani az adatbázisban található feature vector-okkal. Ez első hallásra igen „tudományosan” hangzik, de gyakorlatilag arról van szó, hogy adott $(a_1, a_2, a_3, a_4, \dots a_n)$ számsorozatot, azaz vektort hasonlítunk össze az adatbázisban található $(x_1, x_2, x_3, x_4, \dots x_n)$ sorozatokkal, ahol az egyes számok valamilyen, a képből kinyert leíró (szemek távolsága, homlok magassága stb.) tartalmazznak. Ez szerencsére pofonegyszerű, hiszen vektorok közötti távolságot nagyon egyszerű számolni. A gyengébb rendszerek gyenge pontja legtöbbször a nem megfelelő leírók használata, illetve azok sikertelen és pontatlan detektálása.

A felismerő rendszerek gyakran használnak tanuló algoritmusokat. Ez azt jelenti, hogy a téves és a sikeres felismerésből is tanul a rendszer, módosítja belső „állapotát”, így a továbbiakban egyre jobb eséllyel fog működni. Ezek a rendszerek nem is feltétlenül igényelnek kezdeti adatbázist, helyette be lehet őket tanítani. A tanuló rendszerek lelke legtöbbször egy neurális háló, ami – az agyunkhoz hasonlóan – nem más, mint elemi „kapcsolók”, neuronok igen nagy és összetett hálózata. A hálózat bemenete a feature vector (leírók sorozata), kimenete pedig lehet akár egy, akár több igen/nem kapcsoló. A betanítás szakaszában különböző képeket „mutatunk” a rendszernek, és figyeljük a kimenetet. Ha egy adott hálózati kapcsolat erősítésével sikeresebb a felismerés, akkor erősítjük, ellenkező esetben gyengítjük. Megfelelően nagy számú és tartalmú kép, valamint optimálisan kialakított háló és feature-ök használatával a rendszer képes lesz sikeresen felismerni azokat a képeket is,

amelyek nem szerepeltek az eredeti tanító-halmazban!

Jellemzően akkor érdemes ilyen megközelítést alkalmazni, ha nem teljesen világos, hogy mik a karakterisztikus jegyek, milyen tulajdonságok alapján lehet egyértelműen felismerni a cél-objektumot. Ennek megfelelően arc, rendszám vagy ujjlenyomat felismerésre nem annyira, ellenben orvosi és ipari röntgenfelvételek (utóbbi források ellenőrzésére) elemzésére sikeresen alkalmazhatók. Természetesen nagyon körültekintően kell alkalmazni ezeket a tanuló algoritmusokat, hiszen sok olyan tényező van, ami megzavarhatja a működést. A példa kedvéért említsük meg a „túltanulás” iskolapéldáját: nem kizárt, hogy a neuronális háló tökéletesen megtanulja a teszhalmazban szereplő képeken látható valódi és „hamis” elváltozások sikeres felismerését, és mégis képtelen lesz egy új kép azonosítására. Ennek elkerülésére nemcsak egy tanító halmazt hoznak létre, hanem egy teszhalmazt is, így pontos képet kaphatunk arról, hogy mennyire sikeresen működik a rendszer. Jegyezzük meg, hogy bizonyos alkalmazások nem teszik lehetővé a betanítást, azaz a „robot” élesben tanul.

A felismerés mellett a modellillesztés az egyik legfontosabb probléma a gépi látás területén. Ebben az esetben tudjuk, hogy mi látható a képen, de bizonyos tulajdonságait, paramétereit ki kell találnunk. Például:

- Tudjuk, hogy egy kéz van a képen, de ki kell találnunk, hogy pontosan milyen állásban.
- Tudjuk, hogy egy autó van a képen, de azt nem, hogy milyen irányban áll.
- Tudjuk, hogy az úttest képét látjuk, de nem tudjuk, hogy milyen messze vagyunk a padkától.

Az, hogy tudjuk, mi van a képen, azt jelenti, hogy rendelkezünk egy modellel, amely alapján meg tudjuk mondani, hogy kitalált paraméterek illeszkednek-e a mérésekhez, vagy nem. Az autót vezető számítógép például tudja, hogy körülbelül hogy néz ki az útpadka az autó bármely helyzetében, és ezt össze is tudja hasonlítani a mérésekkel (azaz a kocsi orrára szerelt kamera képével). Ezek a feladatok is jellemző leírók detektálásával kezdődnek, az ismeretlen paraméterek kinyerése pedig általában két lépésben történik. Az első lépésben, ha mód van rá, speciális szabályok alapján megmondjuk, hogy körülbelül mik a helyes

paraméterek: nagyjából merre van a keresett kéz, hozzávetőlegesen merre áll az autó stb. A második lépésben standard matematikai eszközökkel finomítjuk ezt a megoldást, azaz megkeressük azokat a paramétereket, melyek minimalizálják a különbséget a mért jellemzők, és a modell között. A modellillesztés halmazába tartozik a videó alapú 3D szkennelés és a kameraillesztés is! Itt két dolgot tudhatunk: azt, hogy „merev” felületeket látunk, valamint azt, hogy a felvételt egy perspektivikus leképezést megvalósító kamerával készítettük. A megtalálandó paraméterek: a kamera helyzete, iránya, a fókusztávolság, valamint a térbeli felületek pontjainak 3D helyzete. A modellillesztés problémájánál is alkalmazhatóak tanuló módszerek.



Bár ipari környezetben már széles körben alkalmaznak „látó” robotokat, a mindennapjainkban ritkán találkozunk ezzel a technológiával. A bútorokat elkerülő robotporszívó, és a gazdáját felismerő robotkutyá egyelőre inkább mosolyt csal az ember arcára. Néhány éven belül azonban változhat a helyzet: már most megjelentek az első „út-elhagyásra” figyelmeztető számítógépek az autókban, és gőzerővel folynak a kutatások a tartalom alapján történő kép- és video-keresés területén is. Remélhetőleg ezek az újítások már közvetlenül is javítani fogják életminőségünket.

Vass Gergő



A címképet a rajta látható Robovie-III fejlesztője, az Advanced Telecommunications Research Institute International (ATR) bocsájtotta rendelkezésünkre.