

Računalni vid

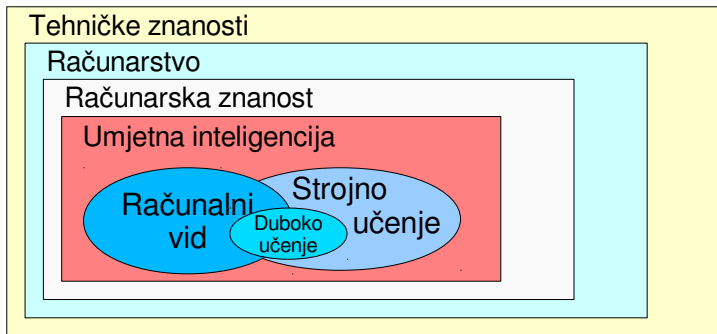
Uvodno predavanje

Siniša Šegvić

UniZg-FER D307

Sveučilište u Zagrebu

O PREDMETU: KARTA POJMOVA



Umjetna inteligencija:

- tehnički sustavi za poslove koje ljudi obavljaju bolje od strojeva

Računalni vid:

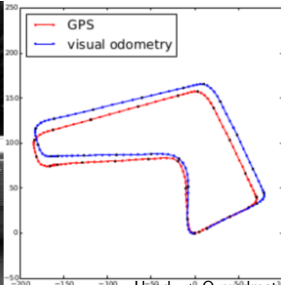
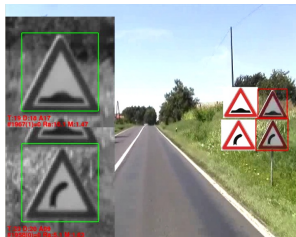
- razumjeti procese strojne percepcije i spoznaje
- stvoriti novu vrijednost kroz napredne usluge

O PREDMETU: RAČUNALNI VID

Proučavamo tehnike za izlučivanje korisnih informacija iz slika

Nekoliko glavnih pristupa prema tom cilju:

- **raspoznavanje**: izlučiti semantičku informaciju: ima li objekta na (x,y) ?
- **rekonstrukcija**: procijeniti geometrijsku informaciju: (ključne točke), (gibanje u slici), položaj kamere, strukturu scene ...



O PREDMETU: METODE

Oba pristupa temelje se na optimizaciji:

- **raspoznavanje**: izlučiti semantiku naučenim modelom
 - kriterij: gubitak na skupu za učenje (+ regularizacija...)
 - optimizacija tijekom učenja, (relativno) jednostavno zaključivanje
 - klasificiranje slika, pronalaženje objekata, segmentacija slika
 - krajnji cilj **kognitivnog** vida: razumjeti sliku
- **rekonstrukcija**: optimirati geometrijski model
 - kriterij: slaganje s mjerenjima u slici
 - klasični pristup: nema učenja, optimizacija tijekom produkcije
 - danas metode raspoznavanja koristimo i za rekonstrukciju

U posljednjem desetljeću svjedočimo ekspanziji **raspoznavanja**:

- ključ: izlučiti kvalitetne slikovne reprezentacije
- danas to provodimo modelima **strojnog učenja**

O PREDMETU: VRIJEME

Sve donedavno, raspoznavanje u "divljini" je bilo akademska disciplina

- primjenljivi rezultati dobivani su samo u laboratorijskim uvjetima
- međutim, vrijedan rad doveo je do fascinantnog napretka

Danas je kognitivni vid tehnologija s jasnom industrijskom vrijednošću

- Google, Facebook, Intel, Microsoft, Nvidia, i mnogi kineski giganti u područje ulažu milijarde dolara
- Google kupio DeepMind for 0.6 B\$ (2014)
- Intel kupio MobileEye for 15.3 B\$ (2017)
- NVidia investirala 20 M\$ in TuSimple (2017)

Pogledajmo zašto je kognitivni vid toliko zanimljiv!

PRIMJENE: AUTONOMNA VOZILA



<https://www.youtube.com/watch?v=9ydhDQaLAqM>

Društvo automobilskih proizvođača definira 6 razina autonomije (2014)

- Teslin autopilot zadovoljava razinu 2 (2016).
- Prototipi razine 4 već su na cestama: Google, Tesla, Uber, Audi, Renault

PRIMJENE: DIJAGNOSTIKA BEZ LIJEČNIKA

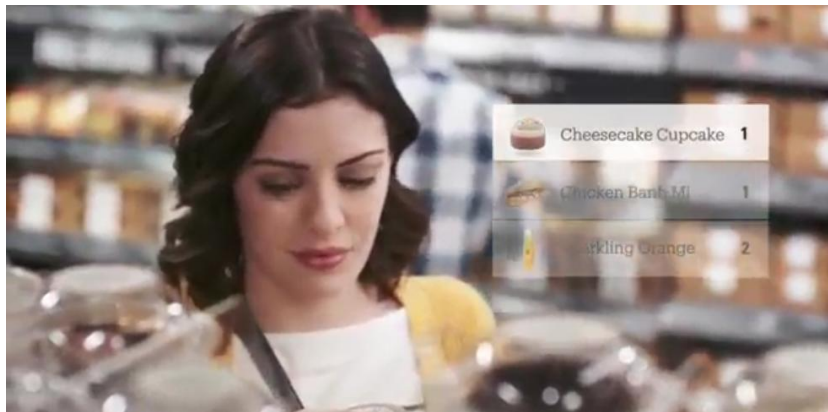


by showing it thousands of pictures of skin conditions.

<http://www.nature.com/nature/journal/v542/n7639/abs/nature21056.html>

Esteva et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 115–118 (02 February 2017)

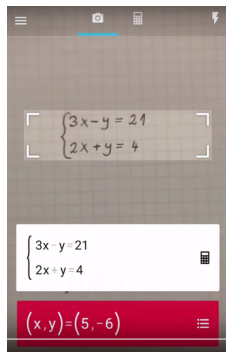
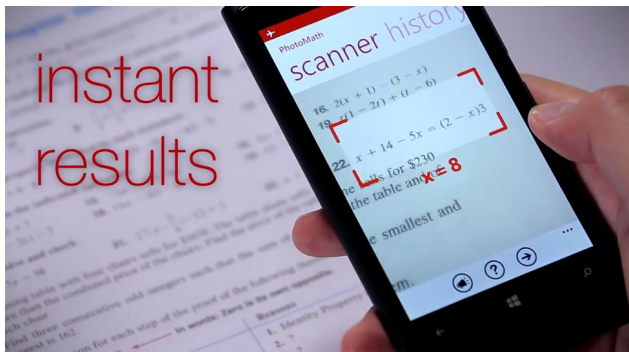
PRIMJENE: DUĆAN BEZ BLAGAJNI



<https://www.youtube.com/watch?v=Jg0pQaV44I8>

Amazon Go, 2131 7th Ave, Seattle, WA.

PRIMJENE: UPLATNICE BEZ ŠALTERA...



<https://www.youtube.com/watch?v=oXn9NuUmhDM>

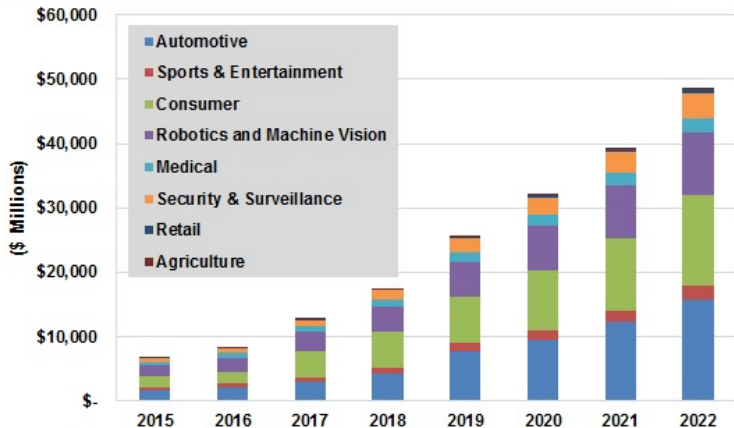
<https://www.youtube.com/watch?v=XIbVB50mlh4>

...i rješavanje matematičkih zadataka (s postupkom!)

PRIMJENE: PROJEKCIJE I PROGNOZE



Computer Vision Revenue by Application Market, World Markets: 2015-2022



Source: Tractica

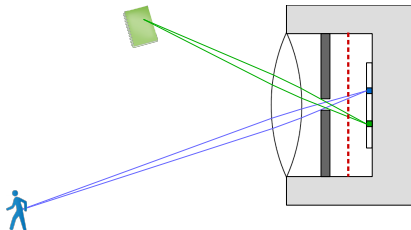
PLAN

1. Uvod i motivacija
 - pristupi strojnog gledanja, primjene
2. Osnove računalnog vida
 - o digitalnim slikama, pogled unatrag
3. Provedbeni plan
 - teme, predavanja, laboratorijske vježbe

DIGITALNE SLIKE: STVARANJE

Digitalna kamera ima četiri glavna elementa:

1. **senzor**: a pravokutno polje osjetilnih elemenata (→ **pixeli!**)
2. **otvor**: rupica koja propušta svjetlost na senzor
 - osjetilne elemente pobuđuju odgovarajući uski snopovi svjetlosti
3. **leća**: uređaj za prikupljanje svjetla (osigurava jači signal)
4. **zatvarač**: osigurava da ekspozicija traje djelić sekunde



Za slike u boji treba nam malo složeniji sustav

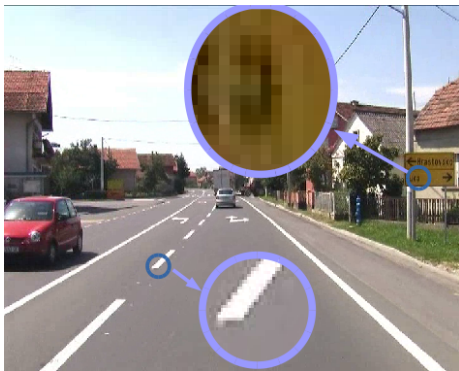
- npr. tri prepletana filtra, tri senzora i sustav ogledala

DIGITALNE SLIKE

Tako, svaka digitalna slika je pravokutno polje piksela

- dimenzije (retci \times stupci) ovise o senzoru

Ako sliku dovoljno povećamo, pikseli postaju vidljivi:

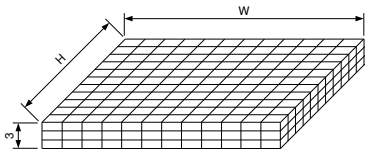


Ako se to ne dogodi, znači da vas program za prikazivanje slika vara :-)

DIGITALNE SLIKE: STRUKTURA

Ako se radi o slici u boji, svaku piknju opisujemo trojkom (R, G, B).

Takve slike predstavljamo tenzorom $H \times W \times 3$:



Ispeglani tenzor možemo smjestiti u memoriju kao polje $H \cdot W \cdot 3$ brojeva:



Sliku od 200×200 piksela možemo predstaviti 120000-D vektorom.

POGLED UNATRAG: POČETAK (1966)

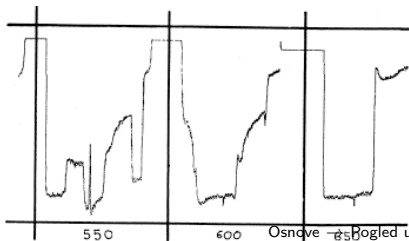
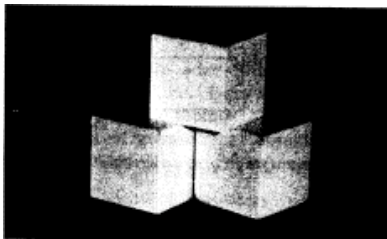
Odmah nakon pojave digitalnih slika postavilo se logično pitanje:

- može li računalo analizirajući sliku razumjeti svijet?

1966 godina: jedan američki profesor predlaže diplomandu da napiše program koji će ispisati što ima u slici.

Taj zadatak je bio ispred svog vremena: problem je mnogo godina ostao neriješen unatoč vrijednom radu mnogih ljudi...

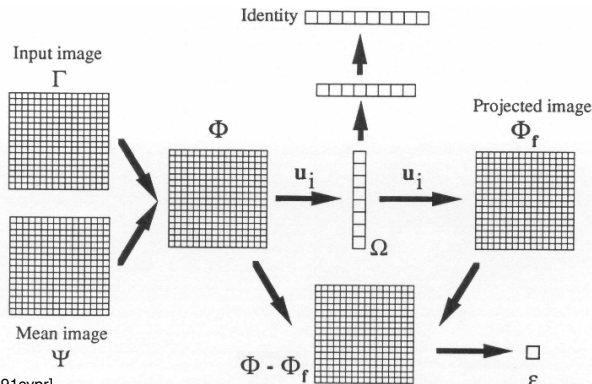
Razumijevanje problema raslo je polako ali ustrajno, ukratko ćemo ponoviti glavna postignuća



POGLED UNATRAG: KLASIFIKACIJA OBJEKATA (1991)

Recept za klasificiranje poravnatih slika lica [turk91cvpr]:

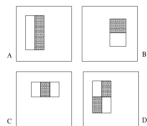
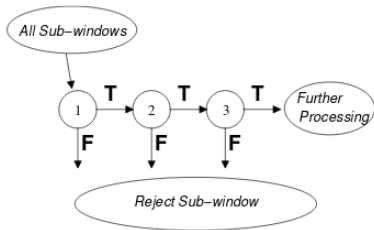
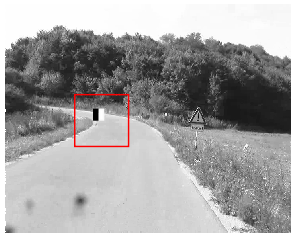
- slike se tretiraju kao visokodimenzionalni podatci
- hypotetizira se da slike nastanjuju nižedimenzionalni **prostor lica**
- uči se linearna projekcija na prostor lica (podatci za učenje!)
- provodi se raspoznavanje na **reprezentaciji slike** u prostoru lica



POGLED UNATRAG: PRONALAZENJE OBJEKATA (2001)

Jednostavni objekti, stvarno vrijeme [viola01cvpr]:

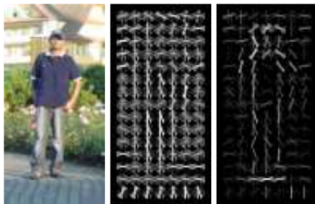
- pronalaženje se tretira kao binarna klasifikacija u **pomičnom oknu**
- klasifikator se uči kao kaskada sve strožih detektora
- pojedine razine kaskade kombiniraju **jednostavne** značajke
- mogućnost klasificiranja preko 100000 okana na 25 Hz



POGLED UNATRAG: PRONALAZENJE OBJEKATA (2005)

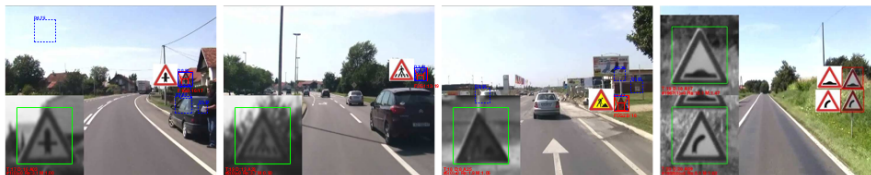
Nešto složeniji objekti [dallal05cvpr]:

- reprezentirati okna ručno dizajniranim lokalnim opisnikom (HOG)
- linearna klasifikacija okana
- moćniji i sporiji od [viola01cvpr]



[dallal05cvpr]

Povezivanje posljednjih dvaju pristupa omogućava pouzdano pronalaženje prometnih znakova:

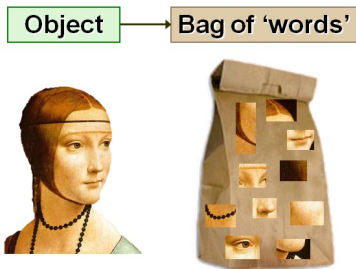
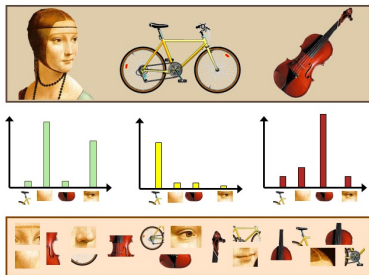


[segvic14mva]

POGLED UNATRAG: KLASIFIKACIJA SLIKA (2004)

Recept za raspoznavanje neporavnatih slika [csurka04eccv]:

- kodirati izgled okana ručno dizajniranim značajkama (SIFT)
- naučiti diskretni rječnik opisnika okana (slikovne riječi)
- reprezentirati slike kao histograme slikovnih riječi
- klasificirati slikovne reprezentacije plitkim klasifikatorom



[fergus09iccv]

MODERNO VRIJEME: JUČER

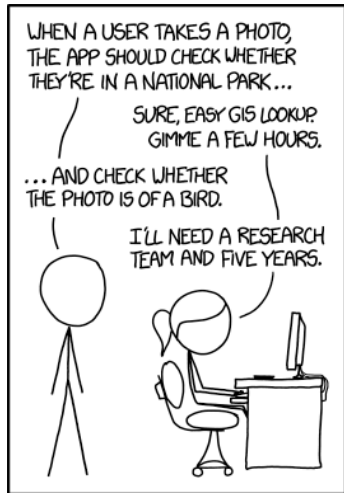
Konačno, svi su saznali da je vid težak...

2014: popularna kultura predstavlja računalni vid kao nemoguću misiju...

- zaključiti sadrži li slika pticu?
- pa, s istraživačkom ekipom i pet godina - možda...

Međutim, od 1966 smo prevalili veliki put:

- pouzdana detekcija ptica moguća još 2010 (s istraživačkom ekipom)
- današnja metodologija i alati omogućuju nam da takve probleme riješimo kod kuće!



IN CS, IT CAN BE HARD TO EXPLAIN THE DIFFERENCE BETWEEN THE EASY AND THE VIRTUALLY IMPOSSIBLE.

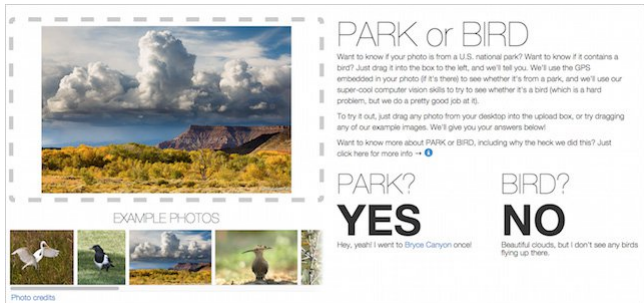
<https://xkcd.com/1425> 24.09.2014

Osnove → Moderno vrijeme 20/25

MODERNO VRIJEME: DANAS

Tako, mjesec dana nakon XKCD 1425, Flickr ponosno javlja:

- riješili smo problem "park or bird"...
- ...u manje od pet godina :-)



The screenshot shows the Flickr 'PARK or BIRD' tool interface. On the left, there is a large photo upload box containing a landscape photo of Bryce Canyon. Below it, a row of five smaller 'EXAMPLE PHOTOS' is shown: two birds, a penguin, a landscape, and a bird. On the right, the title 'PARK or BIRD' is displayed. Below the title, there is explanatory text: 'Want to know if your photo is from a U.S. national park? Want to know if it contains a bird? Just drag it into the box to the left, and we'll tell you. We'll use the GPS embedded in your photo (if it's there) to see whether it's from a park, and we'll use our super-cool computer vision skills to try to see whether it's a bird (which is a hard problem, but we do a pretty good job at it). To try it out, just drag any photo from your desktop into the upload box, or try dragging any of our example images. We'll give you your answers below. Want to know more about PARK or BIRD, including why the heck we did this? Just click here for more info ->'. Below this text are two columns: 'PARK?' with a large 'YES' and the text 'Hey, yeah! I went to Bryce Canyon once!'; and 'BIRD?' with a large 'NO' and the text 'Beautiful clouds, but I don't see any birds flying up there.' At the bottom left, there is a small 'Photo credits' link.

<https://code.flickr.net/2014/10/20/introducing-flickr-park-or-bird/>

Probleme u računarskoj znanosti lako je i potcijeniti i precijeniti :-)

SADRŽAJ: PREDAVANJA

- uvodno predavanje
- geometrijske deformacije slika (lab1)
- detekcija rubova i kuteva (lab2)
- ključne točke u prostoru mjerila
- analiza kretanja u slikovnoj ravnini
- TBD
- TBD
- (međuispit)
- diskriminativni modeli za klasifikaciju slika
- naučena detekcija i segmentacija (lab3)
- generativni modeli slika (lab4)
- transformeri
- TBD
- TBD
- (završni, klasični ispit)

SADRŽAJ: LABORATORIJSKE VJEŽBE

1. geometrijske deformacije slika
2. detekcija rubova i kuteva
3. detekcija objekata
4. normalizirajući tokovi

Neslužbene stranice predmeta:

<https://unizg-fer-d307.github.io/compvis/>

SADRŽAJ: NASTAVNICI

Predavanja:

- Tomislav Hrkać
- Josip Šarić
- Siniša Šegvić
- ...

Vježbe:

- Josip Šarić
- Petra Bevandić
- Matej Grcić
- ...

ZAHVALA

Ova predavanja proizišla su iz istraživanja koje je financirala Hrvatska zaklada za znanost projektom ADEPT.

<http://adept.zemris.fer.hr>