

# Hierarhično iskanje ujemanja za vizualno kategorizacijo objektov

Matjaž Jogan

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, Tržaška 25, 1000 Ljubljana, Slovenija  
E-pošta: matjaz.jogan@fri.uni-lj.si

**Povzetek.** Predstavljamo model vizualne prepoznavne in kategorizacije objektov na podlagi odkrivanja intenzitetnih in strukturnih ujemanj med interpretirano sliko in prototipno sliko objekta. Glavni prispevek te raziskave je uporaba sinhronega hierarhičnega iskanja ujemanj, pri čemer se visokonivojska ujemanja konstruirajo pragmatično, v več ravneh združevanja, izbire in inhibicije. Z iskanjem ujemanj lahko kategoriziramo objekte brez pretirane uporabe učenja, le z izčrpnim iskanjem podobnosti med objekti. Kategorije torej niso opredeljene z naborom značilik, ki bi omogočale optimalno kategorizacijo, temveč kot mrežni sistem podobnosti. Eksperimentalni rezultati dokazujejo primernost metode za iskanje višjenivojskih ujemanj med prototipno sliko objekta v kanoničnem pogledu in sliko neznanega objekta.

**Ključne besede:** računalniški vid, vizualna kategorizacija, vizualno ujemanje, vizualno učenje, hierarhične metode

## A Hierarchical Matching Framework for Visual Object Categorization

**Extended abstract.** We investigate a framework for visual object categorization in artificial cognitive systems that is based on discovery of appearance and structural similarities between prototypes that represent a category and object exemplars. Our main contribution is a novel approach for visual categorization of objects by synchronous hierarchical matching to a prototype, where high-level matches between an object and a prototype are gradually discovered through several steps of binding, selection and inhibition. We show that categorization can be achieved without an excessive collection of evidence or learning from examples. We start with a low-level description, which defines relatively stable local regions of interest (ROI) based on their appearance (Figure 2). In order to characterize the appearance variation within local regions, the system learns a codebook of ICA filters which impose a maximally sparse response (Figure 3). Clusters of local features that conform to a subset of Gestalt rules (being essentially co-centric or co-linear) are then grouped (Figure 4). Features on subsequent levels are then dynamically constructed and matched in synchrony between the view being interpreted and a prototype view of an object. The composite features are augmented with structural properties and represented in a geometric conceptual space, the structure of which is learned on a prototype. Local binding is performed through the hierarchy (Algorithm 1) until a focused response within an area that encompasses the whole object is constructed (Figures 5 and 6). The result is a set of matches of composite features at level  $N$ , where each of the features can be tracked to the constituent features at levels  $N - 1, N - 2, \dots, 0$ . The decision whether the two object categories match, can be done based on the number of high-level matches. The experiments on the ETH-80 database show that the method efficiently discriminates between eight categories of objects in a canonical view.

**Key words:** computer vision, visual categorization, visual matching, visual learning, hierarchical methods

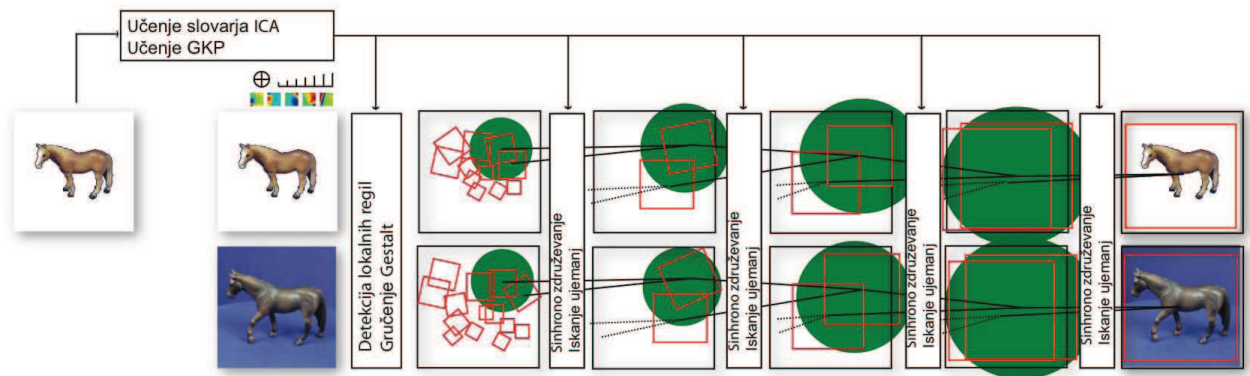
Prejet 5. februar, 2009  
Odobren 14. februar, 2009

---

### 1 Uvod

Kategorizacija objektov spada med pomembnejše probleme umetnega zaznavanja. V tem delu obravnavamo ožje področje vizualne kategorizacije objektov, ki proučuje problem predstavitve in prepoznavanja kategorij objektov z uporabo vidne informacije. V okviru razvoja umetnih kognitivnih sistemov je vizualna kategorizacija pomembna predvsem kot funkcionalnost, ki premosti razliko med procesiranjem signala in višjenivojskim semantičnim procesiranjem, saj omogoča abstraktno interpretacijo prizorov in situacij.

Eden prvih in vplivnejših poskusov računskega modela za kategorizacijo objektov je Marrov model [12], ki temelji na postopnem procesu prepoznavne tridimenzionalnih primitivov kot osnovnih gradnikov, iz katerih so sestavljeni kompleksnejši objekti. Marr predvideva, da lahko primitive uspešno rekonstruiramo, nakar prizor kategoriziramo s pomočjo modela, na primer z uporabo interpretacijskih dreves [8]. Težavna rekonstrukcija primitivov in omejitve primitivov pri predstavitvi nestrukturiranih objektov so pogojevale postopen odmik od visokonivojskih predstavitev k izvorni slikovni informaciji, npr. k predstavitvam, zasnovanim na globalnem videzu objektov, ali pa k predstavitvam z lokalnimi značilkami. Lokalne predstavitve so še posebej uspešne kot osnova za statistično zasnovane modele kategorizacije. Ti se razli-



Slika 1. Shema modela za hierarhično iskanje ujemanj. S pravokotniki so uokvirjena lokalna področja, krogi pa pomenijo receptivna polja združevanja. Zavoljo preglednosti je na sliki prikazan le del hierarhičnega drevesa povezovanja dveh začetnih lokalnih regij. Figure 1. The outline of the framework for hierarchical matching. Rectangular outlines denote local regions, while circular regions denote binding receptive fields. For the sake of clarity, only one branch of binding is depicted, starting with two initial local regions.

kujejo predvsem po vrsti lokalnih značilnk in po načinu in stopnji integracije geometričnih relacij med značilnkami. Leibe in sod. [9] tako implementirajo kategorizacijo s slovarji lokalnih zaplat in implicitnim geometrijskim modelom, medtem ko Fei-Fei in sod. [2] ter Torralba in sod. [17] uporabljajo generativni model lokalnih konstelacij. Hierarhične predstavitve [4, 3, 14, 6], ki predvidevajo hierarhično kompozicionalnost objektov, poskušajo modelirati dele kot konstelacije lokalnih značilnk na večih nivojih podrobnosti.

Medtem ko našeti pristopi definirajo kategorije z učenjem značilnosti nad množico primerov ali z namensko definicijo lastnosti, pa lahko kategorije opredelimo le na podlagi podobnosti s prototipom [15], kjer podobnost vrednotimo z iskanjem strukturnih ujemanj [16], z deformacijskim ujemanjem oblike [1] ali s primerjavo segmentacijskih dreves [18]. Večina predlaganih metod temelji na predhodnem iskanju lokalnih ujemanj, ki pa je v kontekstu kategoričnega primerjanja objektov pogosto problematično.

V predlaganem pristopu predlagamo model kategorizacije s postopnim iskanjem lokalnih in strukturnih ujemanj s prototipom, kjer se ujemanja konstruirajo v hierarhičnem sosledju lokalnega združevanja in inhibicije. Tak način iskanja omogoča intenzivno odkrivanje podobnosti na več nivojih podrobnosti. Podobnosti, ki naj predmet opredelijo kot pripadnika določene kategorije, so torej izpeljane iz prototipne slike objekta v kanoničnem pogledu [15] in niso vnaprej opredeljene. Učenje kategorije ne predvideva obširnega zbiranja znanja o kategoriji objektov, temveč je implementirano kot sprotno učenje lastnosti prototipa in kot učenje in prilagajanje parametrov združevanja. Diagram na sliki 1 prikazuje poenostavljeno shemo poteka kategorizacije. Prototipna slika je podlaga za učenje lokalnega slovarja filtrov za klasifikacijo lokalnih regij in za učenje lastnosti geometričnega konceptualnega prostora (GKP). Učenju sledi več nivo-

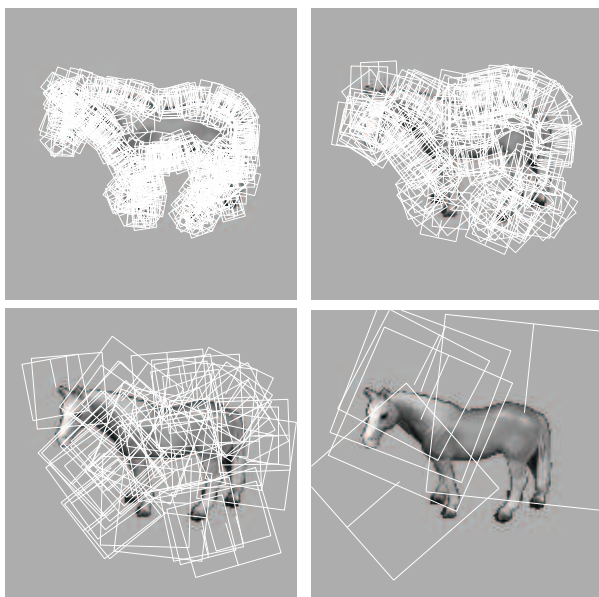
jev združevanja značilnk, iskanja ujemanj in inhibicije. Značilke se združujejo v pare, ki z vsakim nivojem rastejo, tako glede na velikost in kompleksnost sestavljenih značilnic kot tudi glede na površino slike, ki jo opisujejo. Ti procesi potekajo sinhrono, v okviru omejenih receptivnih polj prototipa in interpretirane slike. Če je bilo iskanje ujemanj med prototipom in interpretirano sliko uspešno, lokalna področja na končnem nivoju obsegajo celotno površino prototipa in ustrezno površino interpretirane slike. Število visokonivojskih ujemanj odraža strukturno in vizualno podobnost s prototipom. Če sistem kategorij zasnujemo kot mrežo podobnosti, lahko na podlagi števila ujemanj objekt kategoriziramo.

V nadaljevanju bomo najprej predstavili postopek detekcije, učenja in gručenja lokalnih regij, ki mu bo sledil opis hierarhičnega združevanja. Zaključili bomo z opisom eksperimentalnih rezultatov in kritičnim ovrednotenjem metode.

## 2 Detekcija, učenje in gručenje lokalnih regij

### 2.1 Detekcija

Z detekcijo stabilnih lokalnih regij želimo zagotoviti lokalno podporo za poznejše združevanje slikovne informacije. Klasične metode za razpoznavanje objektov na podlagi lokalnih regij uporabljajo informativne opisnike, ki so načrtovani z namenom minimizacije števila napačnih ujemanj [11]. Pri kategorizaciji objektov pa je značilno, da se objekti iz določene kategorije na lokalni ravni lahko zelo razlikujejo. Zato v pričujočem pristopu lokalno fotometrično informacijo uporabimo le kot osnovo za gradnjo značilnk na višjih nivojih, ki opisno moč pridobijo šele na nivoju posameznega objekta. Kljub temu pa morajo nizkonivojske značilke ustrezati merilom ponovljivosti in stabilnosti, imeti pa morajo tudi lastno merilo in orientacijo. Poleg tega želimo, da so lokalne regije gosto



Slika 2. Lokalne regije  $K_i$ , detektirane na prvih štirih oktavah. Pravokotniki ponazarjajo velikost regij in njihovo usmerjenost. Figure 2. Local regions  $K_i$  detected at the first four octaves. The size and orientation of the regions is denoted by oriented rectangles.

posejane, tako da je začetna predstavitev redundantna.

Lokalna območja poiščemo z detektorjem lokalnih regij, ki je zasnovan na podlagi lastnosti diferenčnih operatorjev v prostoru meril [10, 11]. Sliko  $I(x, y)$  preslikamo v prostor meril  $L(x, y, \sigma)$ , do katerega pridemo s konvolucijo slike z Gaussovimi jedrom  $G(x, y, \sigma)$  spremenljive velikosti. Središča lokalnih regij definiramo kot lokalne ekstremske razlike Gaussovih jeder (Difference of Gaussians, DoG)

$$DoG(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (1)$$

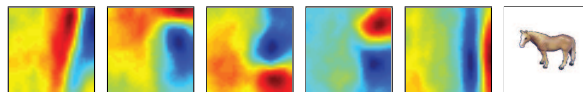
v prostoru meril. Izberemo tiste kandidate, ki presežejo prag lokalnega kontrasta in so lokalni ekstrem tudi v lokalni slikovni soseščini. Kandidate za lokalne regije dodatno okarakteriziramo glede na prisotnost roba v središčni točki.

Velikost posamezne regije je določena z merilom, na katerem je bila regija detektirana. Usmerjenost regije je določena kot smer, kjer histogram lokalnih gradientov zavzame maksimalno vrednost [11].

Lokalne regije  $K_i$  so torej opisane z lokacijo  $^cK_i$ , usmerjenostjo  $^oK_i$  ter velikostjo  $^rK_i$ . Slika 2 prikazuje regije, detektirane na prvih štirih oktavah.

## 2.2 Učenje slovarja ICA

Lokalne regije so zdaj opisane le s strukturnimi značilnostmi regije. Opis videza regije podamo na podlagi slovarja filtrov, katerih maksimalno razpršeni odzivi določajo razred lokalne regije. Število razredov in s tem



Slika 3. Slovar ICA, naučen na prototipni sliki. Figure 3. ICA codebook learned from a prototype image.

velikost slovarja je namenoma majhno, saj se želimo izogniti natančnemu opisu, ki bi preprečilo konstrukcijo visokonivojskih ujemanj.

Nabora filtrov se naučimo z analizo neodvisnih komponent (*Independent Component Analysis, ICA*) slikovne matrike učnih regij, ki jih pridobimo iz prototipne slike, ali pa iz več slik objektov določene kategorije. V literaturi se slovarji ICA pogosto obravnavajo kot model nevronskega receptorja v primarnem vidnem korteksu [13, 5], vendar pa učenje slovarja vedno temelji na naključno izbranih regijah naključnih slik. V nasprotju s pristopi v literaturi v tem delu računamo slovar le na informiranih področjih, ki so bili že izbrani v postopku detekcije. Slovar se zato izogne modeliranju redundantne informacije in se osredini na opise značilnih struktur z generično smerjo in velikostjo, ki je določena že z algoritmom detekcije. Če poleg tega učenje omejimo na predmete ene kategorije, lahko že manjši slovar opiše značilne dele objektov (slika 3).

Slikovno matriko  $X$  sestavimo iz normaliziranih intenzivnih vektorjev  $\mathbf{x}$ , ki opisujejo posamezne regije v generični smeri. Posamezen vektor  $\mathbf{x}_i$  lahko predstavimo z linearno superpozicijo več izvorov  $\mathbf{b}_j$ ,

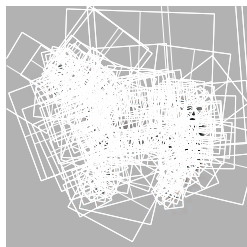
$$\mathbf{x}_i = \sum_{j=1}^N a_{ij} \mathbf{b}_j \quad (2)$$

uteženo s koeficienti  $a_{ij}$ . Matrično obliko enačbe 2 zapišemo kot  $\mathbf{x} = \mathbf{A}\mathbf{b}$ . Če z  $\mathbf{u}$  označimo izvore, ki jih rekonstruiramo na podlagi opazovanja vidnih vzorcev, potem preslikavo  $\mathbf{x}$  v  $\mathbf{u}$  zapišemo kot  $\mathbf{u} = \mathbf{W}\mathbf{x}$ , kjer je  $\mathbf{W} = \mathbf{A}^{-1}$ , če je  $\mathbf{A}$  invertibilen linearni sistem. Z analizo neodvisnih komponent pa izračunamo  $\mathbf{W}$  in  $\mathbf{b}$  tako, da pogojujemo statistično neodvisnost izvorov  $b_j$ . Za izračun slovarja uporabimo algoritem FastICA [5]. Slika 3 je primer slovarja petih filtrov po učenju iz prototipne slike kategorije. Lokalne regije  $K_i$  dodelimo razredu  ${}^wK_i = m$ , ki je določen z indeksom filtra  $w_m$  z najvišjo absolutno vrednostjo odziva

$${}^wK_i = \arg \max_m |\hat{u}_m|; \quad \hat{\mathbf{u}} = \widehat{\mathbf{W}} \hat{\mathbf{x}}. \quad (3)$$

## 2.3 Gručenje Gestalt

Lokalne značilke, ki so osredinjene na robnih točkah, tipično nastopajo v redundantnih, gostih gručah. Redundantnost zmanjšamo z gručenjem robnih regij na podlagi principov *Gestalt*. Algoritem implementira princip



Slika 4. Značilke  $H^0$ , pridobljene po gručenju  $K_i$ .  
Figure 4.  $H^0$  features after grouping of  $K_i$ .

bližine (gručijo se le značilke v omejenem območju), podobnosti (gručijo se le značilke, ki pripadajo istemu razredu) in skupne usode (gručijo se le značilke, katerih smeri tvorijo kongruenten kot  $\alpha$  radianov). Slika 4 prikazuje rezultat gručenja in pridobljene značilke. Na podlagi gručenja so nove značilke  $H^0$  izpeljane iz lastnosti gruči  $K$  tako, da podedujejo razred odziva na slovar ICA  ${}^w H^0 = {}^w K_i$ , smer in velikost značilke pa se izračunata iz geometrijskih lastnosti gruče.

## 2.4 Hierarhično združevanje

Združevanje značilk poteka prek več hierarhičnih nivojev. Značilke postopoma pridobivajo dodatno informacijo o relativni geometrijski konfiguraciji. Ker so geometrijske relacije podane relativno in na lokalni ravni, so značilke neodvisne od globalnega koordinatnega sistema, lokalni proces združevanja pa je popolnoma neodvisen od drugih procesov. Proces združevanja se začne s korakom  $H^0 \rightarrow H^1$  in se načeloma lahko izvede za poljubno število nivojev. Na vsakem nivoju se  $n$  značilk v lokalnem območju združi v  $n$ -terico. Ker združevanje večjega števila značilk hitro privede do prevelike informativnosti značilk, hierarhija združuje značilke le v pare.

Par značilk uredimo glede na velikost regij,  ${}^r H_i^k > {}^r H_j^k$  ter ga opišemo z atributi relativne velikosti  ${}^s H^{k+1}$ , razdalje  ${}^l H^{k+1}$ , kotov  ${}^{\alpha_1} H^{k+1}$  in  ${}^{\alpha_2} H^{k+1}$  med smerjo regij in povezovalno daljico in z oznako medsebojne lege smernih vektorjev  ${}^{or} H^{k+1}$ :

$$\begin{aligned} {}^s H^{k+1} &= \log \left( 1 + \frac{{}^r H_i^k}{{}^r H_j^k} \right) \\ {}^l H^{k+1} &= \log \left( 1 + \frac{\| {}^c H_i^k - {}^c H_j^k \|_2}{{}^r H_i^k} \right) \\ {}^{\alpha_1} H^{k+1} &= \angle ({}^o H_i^k, \overline{{}^c H_i^k - {}^c H_j^k}) \\ {}^{\alpha_2} H^{k+1} &= \angle ({}^o H_j^k, \overline{{}^c H_i^k - {}^c H_j^k}) \\ {}^{or} H^{k+1} &\in \{0, 1, 2, 3\} \end{aligned}$$

Atribute predstavimo v diskretiziranem geometričnem konceptualnem prostoru (GKP), ki ustreza psihološkimi modelom zaznavanja [7]. Geometrične attribute para  $H_i^k$

**Input:**  $\hat{H}^k, H^k, F^k, inner, outer, f$   
**Output:**  $\hat{H}^{k+1}, H^{k+1}, F^{k+1}$   
**foreach**  $F_u^k = \{(\hat{H}_i^k, H_j^k)\}$  **do**  
     $rMin = {}^r H_i^k * inner$   
     $rMax = {}^r H_i^k * outer$   
     $\hat{A} = \{\hat{H}_l^k : {}^c \hat{H}_l^k \in \widehat{RF}(rMin, rMax)\}$   
    **while**  $card(\hat{A}) < f \wedge rMin > 0$  **do**  
        **if** *soseščina znotraj slike* **then**  
             $rMax = rMax + eps$   
        **else**  
             $rMin = rMin - eps$   
        **update**  $(\hat{A})$   
     $\hat{C}^{k+1} = bind(\hat{A}, \hat{H}_i^k)$   
     $\widehat{RF} = convexHull(\hat{A})$   
    preslikaj  $\widehat{RF} \vee RF$   
     $A = \{H_m^k : ({}^c H_m^k \in RF) \wedge$   
         $(\exists \hat{H}_n^k \in \hat{A} : (\hat{H}_n^k, H_m^k) \in F^k)\}$   
     $C^{k+1} = bind(A, H_j^k)$   
     $F^{k+1} = F^{k+1} \cup match(\hat{C}, C)$   
     $\hat{H}^{k+1} = \hat{H}^{k+1} \cup \{\hat{C}_m^{k+1} : \hat{C}_m^{k+1} \in F^{k+1}\}$   
     $H^{k+1} = H^{k+1} \cup \{C_m^{k+1} : C_m^{k+1} \in F^{k+1}\}$

**Algoritem 1:** Hierarhično združevanje, iskanje ujemanj in inhibicija.

**Algorithm 1:** Hierarchical binding, matching and inhibition.

opišemo z diskretnim opisnikom  ${}^G H_i^k$ :

$$\left[ Q_s({}^s H_i^k), Q_l({}^l H_i^k), Q_{\alpha}({}^{\alpha_1} H_i^k), Q_{\alpha}({}^{\alpha_2} H_i^k), {}^{or} H_i^k \right],$$

kjer  $Q(\bullet)$  pomeni kvantizacijo. Naj  $H^k \rightarrow H^{k+1}$  označuje nivoje združevanja. Značilke na vsakem od nivojev lahko opišemo z diskretnim zaporedjem, ki popolnoma opisuje podrejeni značilki ter geometrijska razmerja med njima:

$${}^D H_i^0 = {}^w H_i^0 \quad (4)$$

$${}^D H_m^{k+1} = {}^G H_m^{k+1} \parallel {}^D H_i^k \parallel {}^D H_j^k; k > 0. \quad (5)$$

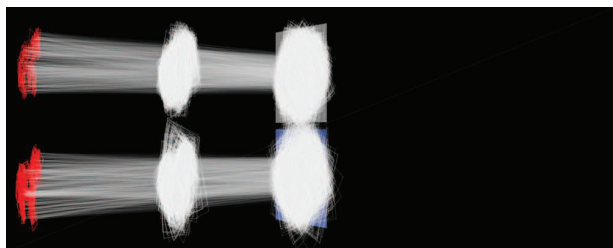
Ker so distribucije geometrijskih parametrov na posameznih nivojih hierarhije odvisne od strukturnih lastnosti objektov, lahko kvantizacijo prilagodimo posameznemu prototipnemu modelu. Model pridobimo s hierarhičnim združevanjem prototipa s samim seboj. Tako pridobljene distribucije parametrov so unimodalne, zato lahko kvantizacijo prilagodimo na podlagi statističnih momentov.

## 2.5 Sinhrono združevanje, iskanje ujemanj in inhibicija

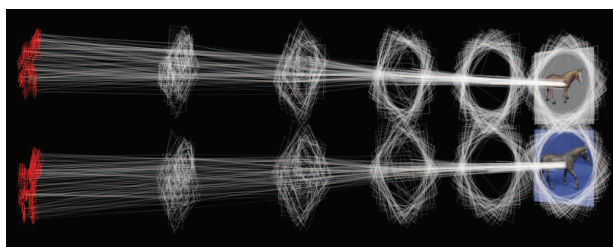
Naj bodo  $\hat{H}^k$  značilke prototipa,  $H^k$  značilke interpretirane slike,  $F^k$  pa indeks ujemanj

$$F^k = \left\{ (\hat{H}_i^k, H_j^k) : {}^D \hat{H}_i^k = {}^D H_j^k \right\}.$$

Združevanje poteka tako, da za vsak par iz množice ujemanj  $F^k$  aktiviramo receptivno polje  $\hat{H}_i^k$ , v katerem



Slika 5. Sinhrono združevanje  $H^1 \rightarrow H^2$ .  
Figure 5. Synchronous matching  $H^1 \rightarrow H^2$ .



Slika 6. Sinhrono združevanje  $H^4 \rightarrow H^5$ .  
Figure 6. Synchronous matching  $H^4 \rightarrow H^5$ .

poiščemo  $f$  kandidatov. Del receptivnega polja, ki vsebuje te kandidate, nato preslikamo na nivo  $k$  interpretirane slike, kjer prav tako poiščemo in združimo kandidate za ujemanja. Lokalna ujemanja na  $H^{k+1}$  ohranimo, druge značilke pa zavrzemo. Postopek opisuje algoritem 1. Izhod iz algoritma so značilke  $\hat{H}^{k+1}$ ,  $H^{k+1}$  ter indeks ujemanj  $F^{k+1}$ . Slika 5 ponazarja združene značilke po prvem koraku  $H^0 \rightarrow H^1$ , slika 6 pa ponazarja končni rezultat, če predvidevamo pet nivojev združevanja.

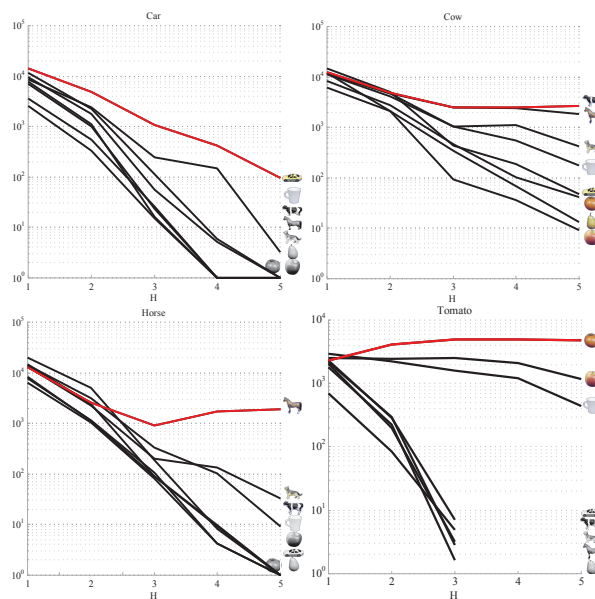
### 3 Eksperimentalni rezultati

Model smo preizkusili na problemu kategorizacije segmentiranih objektov v zbirki *ETH80*. Vsaka od osmih kategorij je predstavljena s prototipnim objektom v kanonični orientaciji. Zaradi manjšega števila detektiranih lokalnih regij učimo slovar ICA filtrov na več slikah iste kategorije, ki pa niso hkrati tudi v testni množici slik. Za vsakega od prototipov se naučimo kvantizacije geometričnega konceptualnega prostora.

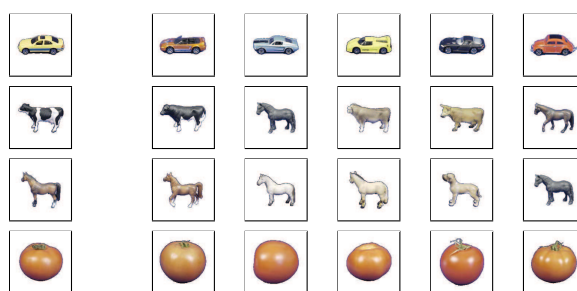
Slika 7 prikazuje povprečno število ujemanj vsakega od prototipov z vsemi drugimi slikami objektov v isti kanonični orientaciji na vsakem od petih nivojev hierarhije. Na desni strani grafikonov so prikazani prototipi kategorij, urejeni po padajočem številu ujemanj na  $H^5$ . Vsi prototipi generirajo najvišje število ujemanj z objekti lastne kategorije, medtem ko dobijo tudi strukturno zelo podobne kategorije v povprečju manj glasov. Iskanje ujemanj s preprostimi predmeti se konča z bistveno manjšim

številom ujemanj pri kompleksnih objektih, saj naučen geometrični konceptualni prostor prilagodi iskanju preprostih oblik, kar tudi do določene mere preprečuje "haluciniranje" preprostih oblik v kompleksnih objektih.

Slika 8 prikazuje prvih pet objektov z najvišjim številom ujemanj na  $H^5$ . Vidimo lahko, da ujemanja na višjih nivojih hierarhije opisujejo konsistentne strukturne podobnosti med objekti ter da so objekti iste kategorije podobenoma med tistimi z največjim številom ujemanj na  $H^5$ .



Slika 7. Povprečna števila ujemanj na nivojih  $H^1 - H^5$ .  
Figure 7. Average number of matches at levels  $H^1 - H^5$ .



Slika 8. Prvih pet objektov po številu  $H^5$  ujemanj.  
Figure 8. The five exemplars with the highest numbers of  $H^5$  matches.

### 4 Sklep

Predstavili smo nov model kategorizacije objektov s postopnim iskanjem lokalnih in strukturnih ujemanj s prototipom, kjer se ujemanja konstruirajo v hierarhičnem sledju lokalnega združevanja, iskanja ujemanj in inhibicije. Rezultati dokazujejo uporabnost metode za kategorizacijo objektov v kanoničnem pogledu. Pomembno je

tudi, da metodo lahko uporabimo za detekcijo potencialnih ujemanj med objekti na več nivojih podrobnosti, kar dosežemo z analizo nizkonivojskih ujemanj, na katerih temeljijo visokonivojska ujemanja. Pokazali smo tudi, da lahko objekte na slikah učinkovito kategoriziramo z uporabo prototipne predstavitve, ki ne zahteva ekstenzivnega učenja na obsežnih učnih množicah.

Metoda omogoča le prepoznavanje v kanoničnem pogledu, vendar pa je bila razvita za implementacijo v širšem kognitivnem sistemu, kjer bi lahko z uporabo kontekstnega procesiranja, večmodalnosti in aktivnega zaznavanja dosegli učinkovito kategorizacijo poljubnih objektov.

## 5 Literatura

- [1] S. Belongie, J. Malik, & J. Puzicha, Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(4):509–522, 2002.
- [2] L. Fei-Fei, R. Fergus, & P. Perona, Learning generative visual models from few training examples. *In Computer Vision and Pattern Recognition Workshop, 2004 Conference on*, pp. 178–178, 2004.
- [3] S. Fidler, M. Boben, & A. Leonardis, Similarity-based cross-layered hierarchical representation for object categorization. *In CVPR*, 2008.
- [4] S. Fidler, & A. Leonardis, Towards scalable representations of object categories: Learning a hierarchy of parts. *In Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pp. 1–8, 2007.
- [5] A. Hyvarinen & P. Hoyer, Emergence of complex cell properties by decomposition of natural images into independent feature subspaces. *In Artificial Neural Networks, 1999. ICANN 99. Ninth International Conference on (Conf. Publ. No. 470)*, Vol. 1, pp. 257–262vol.1, 1999.
- [6] Y. Jin & S. Geman, Context and hierarchy in a probabilistic image model. *In Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, Vol. 2, pp. 2145–2152, 2006.
- [7] G. A. Kelly, *The Psychology of Personal Constructs*. Routledge, 1991.
- [8] J. Krivic & F. Solina, Part-level object recognition using superquadrics, *Computer Vision and Image Understanding*, 95(1):105–126, 2004.
- [9] B. Leibe, A. Leonardis, & B. Schiele, Robust object detection with interleaved categorization and segmentation. *International Journal of Computer Vision*, 77(1-3):259–289, 2008.
- [10] T. Lindeberg & J. Eklundh, Scale detection and region extraction from a scale-space primal sketch. *In Computer Vision, 1990. Proceedings, Third International Conference on*, pp. 416–426, 1990.
- [11] D. Lowe, Object recognition from local scale-invariant features. *In Proc. of the International Conference on Computer Vision, Corfu*, IEEE Computer Society, 1999.
- [12] D. Marr, *Vision*. W. H. Freeman, San Francisco, CA, 1982.
- [13] B. A. Olshausen & D. Field, Emergence of simple–cell receptive field properties by learning a sparse code for natural images. *Nature*, (381):607–609, 1997.
- [14] M. Riesenhuber & T. Poggio, Models of object recognition. *Nature Neuroscience Supplement*, 3:1199–1203, 2000.
- [15] E. Rosch, Natural categories. *Cognitive Psychology*, 4:328–350, 1973.
- [16] A. Shokoufandeh, *et. al*, The representation and matching of categorical shape. *Computer Vision and Image Understanding*, 103(2):139–154, 2006.
- [17] E. Sudderth, A. Torralba, W. Freeman, & A. Willsky, Learning hierarchical models of scenes, objects, and parts. *In Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, Vol. 2, pp. 1331–1338Vol.2, 2005.
- [18] S. Todorovic & N. Ahuja, Extracting subimages of an unknown category from a set of images. *In Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, Vol. 1, pp. 927–934, 2006.

**Matjaž Jogan** je asistent na Fakulteti za računalništvo in informatiko v Ljubljani. Raziskovalno se ukvarja s hierarhičnimi modeli procesiranja vidne informacije. Je avtor in soavtor številnih raziskav s področij umetne zaznave prostorske informacije, samodejnega kartografiranja za lokalizacijo in navigacijo mobilnih sistemov, robustnega in sprotnega učenja, ukvarja pa se tudi z uporabo umetnega vida v interaktivnih sistemih.