Záró kutatói jelentés

Szlávik Zoltán

OTKA, 2011

A kutatás első évében a hangsúly a további kutatást megalapozó tevékenységek elvégzésén volt: eszközök beszerzése, annotációs szoftver létrehozása, képfeldolgozó algoritmusok megvalósítása.

A képfeldolgozó eljárások egységes használatát elősegítendő elkészült egy a Bag of Words modellre épülő keretrendszer. A Bag of Words modell lényege, hogy képi primitívek halmaza jellemzi a képet. A modell kulcseleme a képi primitívek - vizuális szavak -, szótára, mely az összes képi primitív, egyfajta csoportosítással képzett (pl. k-means klaszterezés, hierarchikus k-means, gaussian mixture model), alacsonydimenziós reprezentációja. A modellben a vizuális szavak a képi tartalmat leíró vizuális nyelv alapelemei. A képek ezen alapelemek halmazából épülnek fel. A képleíró vektorok (hisztogrammok) elemei azt reprezentálják, hogy az egyes vizuális szavak hányszor fordulnak elő az adott képen. A fentieknek megfelelően a rendszer 4 fő részből áll, melyek a feldolgozandó képek előfeldolgozásáért és normalizálásáért; a képi jellemzők kinyeréséért és leírásáért; képi jellemzők csoportosításáért valamint a képleíró vektorok előállításáért felelnek. A képi jellemzők detekciója a Laplace of Gaussian, Harris-Laplace, Difference of Gaussian, Harris, MSER (maximally stable extremal region) eljárásokkal lehetséges. A jellemzők leírása SIFT (scale invariant feature transform) illetve HOG (histogram of gradients) leírókkal történik. A leírók összehasonlítása történhet egyszerűbb metrikákkal (L1, L2, chi2 távolság) ill. olyan szofisztikált eljárásokkal, mint az EMD (earth mover distance) távolság és a Pyramid Match Kernel. Ezen jellemzők és leírók alkalmasak a képi tartalom lokális jellemzésére.

A rendszer kezdeti állapotában a képek kategorizálása/klasszifikálása a képleírók (hisztogrammok) direkt összehasonlításával; az SVM (support vector machine), mint bináris klasszifikációs eljárás felhasználásával; vagy a tématanuló PLSA (probabilistic latent semantic analysis) generatív valószínűségi modell, mely az egyes témák (osztályok) együttes előfordulását modellezi, segítségével lehetséges. A kereten belül tesztelhető a különböző jellemzők és leírók hatása a kategorizálás/klasszifikáció pontosságára. A kezdeti kísérletek visszaigazolták a hasonló kategorizáló rendszerekkel kapott eredményeket. A tapasztalataink szerint hatékonyabb kategorizálás érhető el, ha egyszerű detektorokat és nem túl szofisztikált leírókat használunk. Ennek magyarázata az lehet, hogy a túlságosan diszkriminatív képi jellemzők nem az elvárt módon reprezentálják a kép egészét, hanem csak a jellemzők által preferált lokális képi tartalomra fókuszálnak. Ezért az első időszakban megvalósított képi jellemzők és leírók ki lettek egészítve a következőkkel. Képi objektumok jellemzőjeként a lokális mikrostruktúrát valamint a színi információt is megtartó térbeli szűrőcsaládok lettek szoftveresen megvalósítva: MR8, WCM; valamint a globális struktúrát leíró Edgel (éldarabok halmaza) jellemző. A szűrők kimenetét klaszterezve (k-means algoritmussal) kapjuk a textonokat, melyek voltaképpen a bemeneti kép egyfajta textúra- és szín-alapú szegmentálását jelentik. Az Edgel jellemzők robosztus összehasonlítására elkészült a Multi-Scale Oriented Chamfer Matching távolság. Ezen jellemzőket a Medoid-Shift algoritmussal klaszterezve kapjuk az adott osztályra vagy osztályokra legjellemzőbb Edgel jellemzőket.

A felismerés hatékonyságát tesztelve különböző adatbázisokon megállapítottuk, hogy a klasszifikációs eljárások (SVM) hatékonyabbak, mint a generatív modellek (PLSA). Az SVM eljárás hátránya, hogy un. bináris klasszifikációs eljárás és ha több osztályt szeretnénk egyidejűleg osztályozni, akkor több SVM alkalmazására van szükség, ami jelentősen növeli a számításigényt. Ezért úgy döntöttünk, hogy a továbbiakban az SVM eljárásról áttérünk a Boosting eljárások alkalmazására a képi objektumok látványának és struktúrájának a modellezésére. A Boosting eljárásban az osztályozó

függvény gyenge osztályozók súlyozott összege. Az eljárás legnagyobb előnye, hogy hatékonyabban használható többosztályos osztályozási feladatokban mint más bináris osztályozók. A Boosting eljárás gyenge klasszifikátora pl. a Texton jellemzők esetén arra ad választ, hogy adott téglalapon belül egy adott Texton aránya jellemző-e az adott osztályra. Egy ilyen jellemző a lokális látványt jellemzi, összességében pedig ezen jellemzők halmaza leírja hogyan néz ki az adott képi objektum. Az Edgel jellemzők esetében a gyenge klasszifikátor azt vizsgálja, hogy adott Edgel illeszkedése a bemeneti kép élképéhez jellemző-e az adott osztályra. Vagyis melyik strukturális leíró (Edgel) hol illeszkedik a bemeneti képhez meghatározva a lokális ill. összességükben a képi objektum globális struktúráját. A Boosting algoritmus a fenti esetekben kiválogatja az adott osztályra jellemző lokális látványbeli és strukturális jellemzőket felépítve a képi objektum globális látvány ill. strukturális modelljét. Teszteléskor a kiválasztott gyenge klasszifikátorok súlyozott összegét kell kiszámolni a bemeneti képnek azon pontján ahol ellenőrizni szeretnénk az adott objektum meglétét. Általános esetben a bemeneti képen akárhol és akármilyen méretben lehet a keresett objektum. Ezért a képből egy u. n. multi-scale képpiramist kell létrehozni, mely a bemeneti kép különböző méretű változataiból áll és minden kép minden egyes pontjában el kell végezni a kiértékelést. A képi objektumok látványának és struktúrájának együttes modellezésére a Joint-Boosting algoritmust használtuk. Az algoritmus lényege, hogy egységes formátumú gyenge klasszifikátorok használatával lehetővé teszi különböző képi jellemzők együttes használatát a képi objektumok modellezésekor. Vagyis, ha pl. az egyik jellemző a látványt (pl. Texton jellemző) a másik a struktúrát (pl. Edgel jellemző) modellezi, akkor a Joint-Boosting modellben látvány és struktúra együttesen modellezhető. Tesztelve az Edgel jellemző hatását a kategorizálás/klasszifikáció pontosságára megállapítottuk, hogy a végeredmény nagyban függ attól mennyire pontos maguknak az Edgel jellemzőknek az előállítása a bemeneti képen. Ezért ezt a jellemzőt kiváltottuk az egyszerűbb és éppen ezért robusztusabb HOG jellemzővel.

A kutatás utolsó évében az elkészült Joint-Boosting alapú eljárást HOG, LBP és Texton jellemzőkkel használva alkalmaztuk különböző felismerési feladatokban: több osztályos kategorizálás, objektum al-osztályok osztályozása - pl. arcképek "férfi" és "nő" al-osztályokba való besorolása vagy arcképek érzelmek szerinti besorolása; valamint járműtípusok osztályozása. Az objektumok reprezentációja HOG, LBP és Texton képi jellemzők speciális kombinációjával történik, mely kombinációkat kereszt-validációs optimalizálással határoztuk meg.
Többosztályos osztályozásban szokás megkülönböztetni objektum osztályok (pl. autó, ember stb.) osztályozását valamint objektum al-osztályok osztályozását (pl. járműtípusok felismerése). Az első feladat az egyszerűbb, hiszen minden valószínűség szerint kevés vizuális hasonlóság található a különböző objektum kategóriák között. Ezzel ellentétben a második feladat sokkal nehezebb, hiszen, ha pl. járműtípusokat szeretnénk felismerni, akkor szinte biztos, hogy nagyon sok vizuális részlet szinte azonos lesz a különböző típusú járműveken.
Bizonyítandó a kidolgozott eljárások működőképességét először egy viszonylag egyszerű többosztályos osztályozási és detekciós feladatban vizsgáltuk az eljárásunk teljesítményét. Több osztályos kategorizálás tesztelését a Caltech101 képi adatbázison végeztük. Az adatbázis 101 általános kategória képeit tartalmazza, osztályonként 40-800 képpel. A teszteléshez a bankjegy, laptop és karóra osztályokat választottuk; 30 véletlenszerűen választott kép volt a tanítóminta és a maradék képen történt a tesztelés. Az 1. táblázat tartalmazza az objektum detekció átlagos pontosságát 5 kísérlet eredményeinek átlagaként HOG jellemző és Jont Boosting eljárás alkalmazásának eredményeként. Sikeres detekciók példái az 1. ábrán láthatóak.

1. táblázat Többosztályos osztályozás eredménye a bankjegy, laptop és karóra osztályokon

Osztály	Átlagos pontosság %
bankjegy	98
laptop	83
karóra	97

1. ábra Példák sikeres objektum detekcióra

Arcképek különböző osztályok szerinti besorolását a nehéz feladatok közé szokás sorolni. Ennek oka az, hogy a klasszifikációs algoritmusok a legtöbbször nem az osztályokat tanulják meg, hanem a különböző személyeket. Ezért teszteléskor külön figyelmet kell arra fordítani, hogy a tanító és teszthalmazokban szereplő arcképek különböző személyektől származzanak. Arcképek érzelmek szerinti besorolásához a CK+ adatbázist használtuk [7]. Az adatbázis 593 érzelmi állapotot bemutató képsorozatot tartalmaz, melyek közül 327-ről jelenthető ki nagy bizonyossággal, hogy melyik alap érzelmet mutatják be. Követve [7] útmutatásait kísérletünkben egy kiválasztott személy és a hozzá tartozó képek volt mindig a tesztminta és az összes többi kép a tanítóminta. Kereszt-validációs kísérletek sorozata alapján megállapítottuk, hogy ebben a feladatban a HOG és LBP jellemzők együttes alkalmazása a legcélravezetőbb. Az adatbázisban szereplő összes személyre összesített klasszifikációs eredményeket a 2. táblázat konfúziós mátrixa mutatja. Sikeres osztályozás példái a 2. ábrán láthatóak. Az eljárás részletei valamint a különböző kísérletek eredményei megtalálhatóak az [1] és [4] közleményekben. Összevetve eredményeinket más eljárásokkal [7] megállapítható, hogy eljárásunk teljesítménye lényegesen felülmúlja konkurens eljárások teljesítményét.

2. táblázat Arcképek érzelmek alapján történő besorolása

	méreg	undor	öröm	félelem	szomorúság	meglepettség
méreg	86.36	6.82	0	0	6.82	0
undor	1.72	98.28	0	0	0	0
öröm	0	0	100	0	0	0
félelem	3.7	0	14.81	70.37	3.7	7.41
szomorúság	14.81	3.7	0	3.7	77.78	0
meglepettség	0	1.23	0	0	0	98.77

2. ábra Példák arcképek érzelmek szerinti besorolására

Hasonlóan nehéz feladat az arcképek nemek szerinti osztályozása. Kísérleteinkben a FERET adatbázist használtuk [8]. Tanításhoz 100 darab véletlenszerűen választott férfi és nő arcképét használtuk, és a többi képen végeztük el az osztályozás tesztelését. Az 3. táblázat tartalmazza az objektum detekció átlagos pontosságát 10 kísérlet eredményeinek átlagaként [1]. Sikeres osztályozás példái a 3. ábrán láthatóak. Az érzelem felismeréshez hasonlóan a HOG és LBP jellemzők együttes használata vezetett a legpontosabb osztályozáshoz.

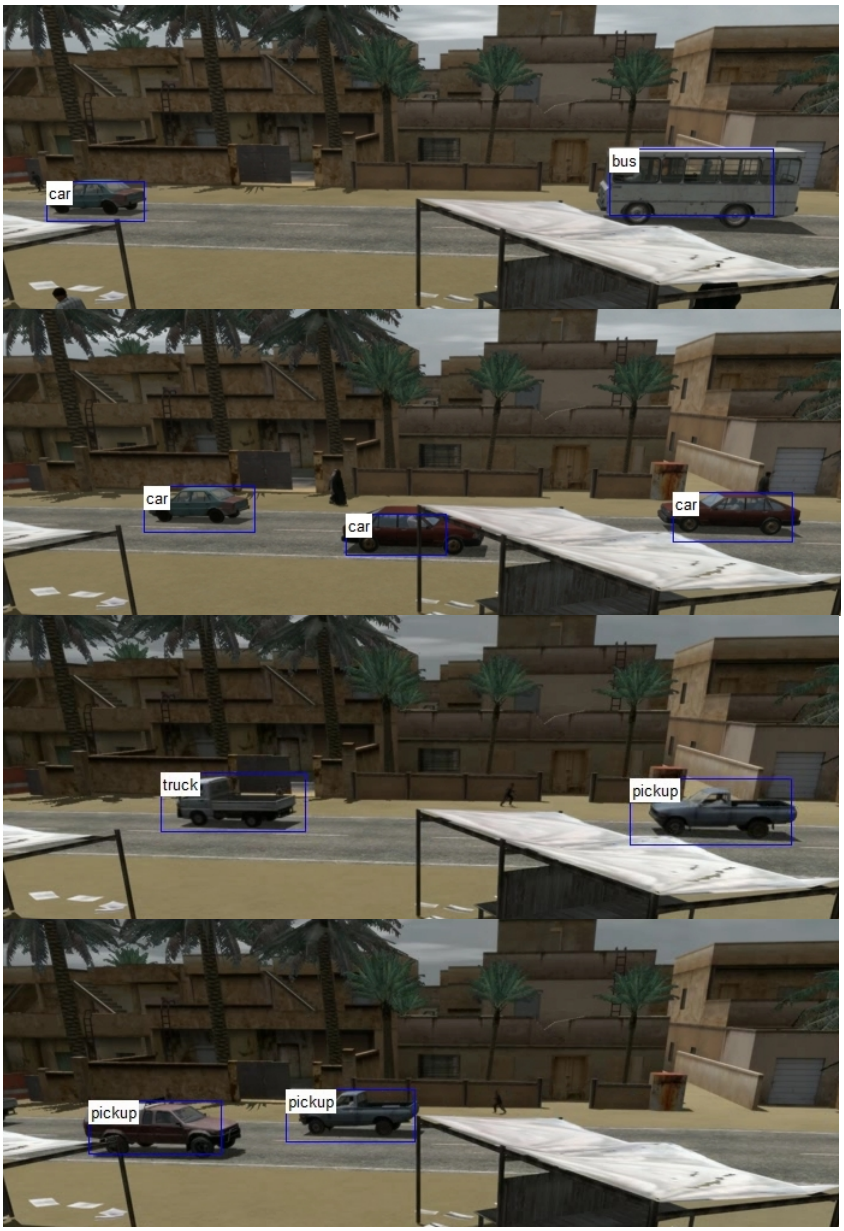
3. táblázat Arcképek nemek szerinti besorolása

Nem	Klasszifikáció pontossága %
Nő	96.5
Férfi	76.5

3. ábra Példák arcképek nemek szerinti besorolására

Legutolsó osztályozási kísérletünkben a kidolgozott eljárást sikeresen alkalmaztuk különböző járműtípusok - személyautó, busz, minibusz, dzsip, teherautó - detekciójára, 4. ábra. A kísérletekhez használt felvételek katonai szimulált környezetből származnak, melyeket manuálisan annotáltunk. Ebben az esetben a Texton jellemzők használata bizonyult a legjobb választásnak. Véleményünk

szerint ennek oka, hogy a szimulált környezetből származó felvételek a valós felvételeknél egyszerűbbek és a képeken megfigyelhető textúrák jobban jellemzik a képi objektumokat, mint a valóságban.



4. ábra Példák járműtípusok sikeres detekciójára

A képi tartalom globális struktúrájának és az objektumok egymáshoz való viszonyának jellemzésének egyik lehetséges módja, ha sikerül őket elhelyezni a valós 3D térben. Ez általában akkor lehetséges, ha egyazon objektumról vagy helyszínről több nézetből áll rendelkezésre felvétel.

Valószínűségi eljárásokat dolgoztunk ki annak meghatározására, hogy a képi objektumok hol érintkezhetnek a föld síkjával a valós 3D térben [2,3,6]. Megmutattuk, hogy több felvétel esetén a képek közötti geometriai transzformáció ismeretében lokális tér- és időbeli statisztikák felhasználásával az objektumok lokalizálhatók a különböző felvételeken lehetővé téve képi információ gyűjtését több kamera nézetből, ezek összehasonlítását és feldolgozását [3,6]. Az eljárás a bemeneti képsorozaton detektált mozgásokról készít időbeni együttmozgási statisztikát. A statisztikákból kiszámolható, hogy a kép mely részein mekkora átlagos méretű objektumok

mozognak. Ezen információ ismeretében pedig a valószínűségi Hough modell segítségével meghatározható az a legvalószínűbb pont, ahol adott időpillanatban az adott objektum a föld síkjával érintkezik. Több nézetből elvégezve a becsléseket és közös paramétertérbe transzformálva, kiátlagolva az eredményeket nagy pontossággal lokalizálhatóak a detektált objektumok. Az eljárás lényege és előnye, hogy csak alacsonyszintű mozgás detekciós eljárást használ és az együttmozgási statisztikákat, vagyis nem használ fel magasabb szintű képi jellemzőket és leírókat.

Eljárást dolgoztunk ki objektumok lokalizációjára többkamerás multimodális környezetben [2]. Egy többkamerás rendszerben a kamerák közötti geometriai kapcsolat ismeretében megadható, hogy a különböző kamerákon észlelt objektumok pozíciója mennyire esik egybe a valós 3D térben. Ha az összehasonlított pozíciók hibahatáron belül vannak, akkor egy objektumhoz tartozónak fogadhatjuk el őket. A kitakarások okozta problémák kezelésére a következő verifikációs eljárást dolgoztuk ki. A megfigyelt objektumok lokális statisztikai jellemzőit a perspektív kontextus [2] szerint modulálva állítjuk elő az objektumok nézet-invariáns képi leíróit. Végül a kapott leírókat elemenként összehasonlítva és valószínűségi térképen aggregálva kapjuk meg az összetartozó objektumokat valamint pozícióikat. Példa sikeres lokalizációra az 5. ábrán láthatóak.

5. ábra Objektumok lokalizációja

A kidolgozott eljárás leglényegesebb tulajdonságai más ismert eljárásokkal szemben, hogy:

- nincs szükség a kamerák kalibrációjára;
- multimodális kamerarendszer esetén is működőképes;
- a megfigyelt objektumok tetszőleges méretűek lehetnek.

A jelenlegi kutatás elsődleges célja vizuális információt reprezentáló modellek kidolgozása volt integrálva a képi objektumok strukturális és megjelenési jellemzőit. A kutatásunk során az objektumok megjelenésének és struktúrájának egy- és több-nézeti modellezésével, különböző vizuális jellemzők egységes modellbe való integrálásával, statisztikai tanulóalgoritmusok alkalmazásával valamint objektumok kategorizálásával foglalkoztunk. A kidolgozott kategorizáló eljárásokat járműtípusok felismerésére valamint arcképek nemek és érzelmek alapján történő osztályozására alkalmaztuk. Az elért eredmények alapján kijelenthető, hogy ezen jellemzők integrálásával jelentősen javítható a klasszikus képi kategorizáló és felismerő algoritmusok hatékonysága. Továbbá, új eljárásokat dolgoztunk ki objektumok lokalizációjára többkamerás multimodális környezetben lehetővé téve képi objektumok több nézőpontbeli összehasonlítását tetszőleges környezetben.

Irodalomjegyzék

1. Zoltán Szlávik, Dömötör Molnár, Joint Boosting of Histogram Like Features for Subject-Independent Facial Image Processing, Pattern Recognition Letters, under review
2. László Havasi, Zoltán Szlávik, A Method for Object Localization in a Multiview Multimodal Camera System, CVPR OTCBVS, 2011

3. László Havasi, Zoltán Szlávik, A statistical method for object localization in multi-camera systems, KÉPAF 2011
4. Dömötör Molnár, Zoltán Szlávik, Joint Boosting of Histogram Like Features for the Generic Recognition of Object Classes and Subclasses, CogInfoCom, 2011
5. László Havasi, Zoltán Szlávik, A statistical method for object localization in multi-camera tracking, IEEE International Conference on Image Processing, ICIP 2010, pp. 3925-3928
6. László Havasi, Zoltán Szlávik, Using location and motion statistics for the localization of moving objects in multiple camera surveillance videos, The Ninth IEEE International Workshop on Visual Surveillance, ICCV 2009, on CD
7. Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I., 2010. The Extended Cohn-Kande Dataset (CK+): A complete facial expression dataset for action unit and emotion-specified expression. Third IEEE Workshop on CVPR for Human Communicative Behavior Analysis.
8. Phillips, P. J., Moon, H., Rizvi, S. A., Rauss, P. J., 2000. The FERET Evaluation Methodology for Face Recognition Algorithms. IEEE Trans. Pattern Analysis and Machine Intelligence. 22, 1090-1104.