

Pázmány Péter Katolikus Egyetem
Roska Tamás Műszaki és Természettudományi Doktori Iskola

Tér-időbeli vizuális jelenségek vizsgálata biológiai alapú algoritmusokkal

Fülöp Tamás

PhD Disszertáció

Témavezető: Zarándy Ákos, DSc



Budapest

2018

Szándékosan üres

Szerintem a kutatás egy körmönfont Sir Arthur Conan Doyle sztori, amiben a válasz ott csücsül a szemünk előtt, csak össze kell rakni a megszerzett ismereteket.

Tartalomjegyzék

1	Bevezetés	2
1.1	Motiváció	4
1.2	A disszertáció felépítése	5
2	Vizuális jelenségek észlelése	6
2.1	Érzékelés és észlelés	6
2.2	Látás, gépi látás	10
2.2.1	A látórendszer működése	10
2.2.1.1	A retina	10
2.2.1.2	A látópályák és feldolgozásuk	18
2.2.2	A gépi látás folyamata	20
2.2.2.1	Látás inspirálta megközelítések	22
2.3	Gyors képfeldolgozás igénye beágyazott rendszerekben	26
2.3.1	Eye-RIS szenzor-processzor tömb	27
2.3.2	Elterjedt megoldások napjainkban	28
3	Objektum közeledésének felismerése biológiai alapú algoritmus segítségével	31
3.1	Közeledő objektum problematikája	31
3.1.1	Térinformáció kinyerése	32
3.1.2	Közeledés a vizuális térben	35
3.1.3	Közeledés érzékelésének klasszikus algoritmikus módszerei	36
3.1.4	Közeledés érzékelésének biológia motiválta módszerei	38
3.1.4.1	Sáskákban található sejtcsoport bemutatása	38
3.1.4.2	Galambokban található sejtcsoport bemutatása	40
3.1.4.3	Az egérben található sejtcsoport bemutatása	41
3.1.4.4	Biológiai modellek összehasonlítása	43
3.2	Biológia motivált közeledő mozgás érzékelés számítógépes modellje	45
3.2.1	Ütközés elkerülési megoldás a sáska látórendszere alapján	45

3.2.2	Egér retina alapú modellezés	45
3.2.2.1	A Pvalb-5 sejtsatorna kvantitatív modellezése	46
3.2.2.2	A modell implementálása	48
3.2.2.3	A modell kalibrációja	52
3.2.2.4	A modell validálása, eredményei	58
3.2.3	A Pvalb-5 alapú modell összehasonlítása más és azóta született biológiai modell eredményekkel	61
3.3	Közeledés irányának meghatározása Pvalb-5 sejt működési modelljének felhasználásával	63
3.3.1	Az eredeti modell módosítása közeledés iránykomponensének számításához	63
3.3.2	A matematikai modell bemutatása	64
3.3.3	A modell implementálása és validációja	66
3.3.4	A modell működése valós környezetben	67
3.4	A biológiai alapú modell lehetőségei és korlátai	67
3.4.1	A Pvalb-5 alapú modell korlátai	68
4	Periodikusan változó fényintenzitású LED-es jelzőlámpa felismerése	69
4.1	Szín alapú szegmentálás	70
4.2	Alul-mintavételezett frekvencia elemzés	71
4.2.1	Goertzel algoritmus	71
4.3	LED jelzőlámpa felismerése Goertzel algoritmus segítségével	74
4.4	Eredmények	77
4.4.1	Összehasonlítás ma elterjedt megoldásokkal	78
4.4.2	A kutatás tárgyát képező modell felhasználásának lehetőségei	80
5	Tudományos eredmények összefoglalása	81
5.1	Kutatás módszere	81
5.2	Új tudományos eredmények	82
5.3	A képfeldolgozási lánc és biológiai vonatkozásai, a kutatás eredményeinek felhasználása, aktualitása	88

6	Publikációs lista	90
7	Köszönet	91
8	Rövidítések jegyzéke	94
9	Függelék	96
9.1	Balra és jobbra történő laterális mozgásnál jelzett irány, valamint mozgó platformon lokális elmozdulás vektorok	96
10	Hivatkozások	97

1 Bevezetés

A világ, amely körülvesz minket, tele van olyan jelenségekkel, melyek egy részét közvetlenül képesek vagyunk érzékelni, míg másokat csak közvetve. Az érzékelés az első lépéscsoja a megismerés, egészen konkrétan az észlelés folyamatának. Az érzékelés minőségén múlik, hogy milyen következtetéseket lehet levonni egy-egy eseményből, mi lesz az észlelés maga.

Az észlelés a biológiai rendszereknél a túlélés egyik nagyon fontos alapkövetelménye. Valamely történés észlelésére adott válasz segíthet elkerülni a sérülést, ütközést, támadást, vagy éppen segíthet a navigációban, táplálék felkutatásában. A percepció feltétlen és feltételes reflexet egyaránt eredményezhet.

A feltételes reflex tanulás alapján jön létre, felfigyelés az összefüggő dolgokra, melyekre aztán cselekvési tervet készít a biológiai szervezet [1]. A feltétlen reflexek esetében nincs tanulás, hanem a biológiai rendszerek eredendően hordozzák azokat az ingerre adott válaszokat, melyek valamely történésre vélhetően optimálisan reagálnak [2]. Az észlelés ebben az esetben nagy jelentőségű, és a törzsfajlódás során folyamatosan lett egyre kifinomultabb. Olyan hatékony mechanizmusokat feltételezhetünk a háttérben, amelyek alapvetően a túlélésre készültek.

A veszélyes szituációk gyors felismerése és elkerülése a mesterségesen létrehozott rendszerek esetén is kritikus feladat. Nem egyszerű válaszolni arra a kérdésre, hogy egy-egy jelenséget milyen jelzésekből kell felismerni, és ehhez milyen megközelítést lehet alkalmazni. A mérnöki intuíciók, a felhalmozott algoritmikus tudás segít a problémákat megoldani, de vannak olyan esetek, amikor a természet kész választ adhat valamely konkrét problémakör megoldására, amely felhasználható az érzékelés kiterjesztésére, az észlelés folyamatának felgyorsítására.

Ilyen biológiailag motivált megoldás lehet, ha kihasználjuk az érzékszervi sajátosságokat, és azt befolyásoljuk, vagy annak működésére hatunk. Ilyen például, ahogy a lakossági gázellátásban a szagtalan metánhoz, szúrós szaganyagot adagolnak, hogy a gázszivárgása azonnal felismerhető legyen. De hasonlóan figyelemfelkeltő egy összetett kezelőpulton az ember figyelmét egy villogó fénnel odacsalogatni, amely a látórendszer változásokat felismerő mechanizmusára hat. A biológiai rendszerekben olyan egyszerű vagy akár összetett reflexek fejlődtek ki, amikre képesek vagyunk hathatni. Ugyanakkor ezeket a

működési mechanizmusokat, ha lehetőség van lemásolni, akkor akár a természetben előforduló alapvető veszélyes szituációkra lehet gyors algoritmikus felismerési megoldásokat adni.

Jelen munkámban a képfeldolgozási algoritmusok biológiai alapú megközelítésével foglalkozom. Olyan megoldásokkal, amelyek segíthetnek bizonyos algoritmus osztályok esetében az eseményfelismerés folyamatát gyorsítani. Különösen aktuális ez manapság, amikor az autókba a legkülönbözőbb vezetést támogató rendszereket szerelik be, amiket az ADAS betűszóval fed le az autóipar, az „Advanced Driver-Assistance Systems” rövidítéseként. A jó minőségű passzív szenzorokra támaszkodó gépi látásnak autóipari jelentőségét az adhatja, hogy az aktív szenzoros érzékeléshez képest – gondolva itt a RADAR-okra és LIDAR-okra egyaránt –, olcsóbbak és egyszerűbb felépítésűek, valamint az aktív szenzorokhoz képest a megnövekedett számú jelforrások zajával sem kell számolniuk. További előnye a képfeldolgozás alapú ADAS megoldásoknak, hogy kiterjeszti az észlelt teret, így a távoli történések észlelésével komolyabb előre tervezésre ad lehetőséget, egyúttal az önvezető autók tényleges megjelenését is már látható távolságba hozta. [3]

Nem mehetünk el azonban azon tény mellett, hogy a gépi látás nem mindig, és nem mindenhol használható önállóan, valamint jellegéből fakadóan örökké hordozni fogja a bizonytalanságot, tekintettel arra, hogy a képek alapján leginkább következtetéseket lehet levonni. A képfeldolgozás alapú objektum- és szituáció-felismerés arra viszont mindenképpen alkalmas, hogy veszélyt jelentő szituációk felismerése az érdekes régiók (Region of Interest, továbbiakban ROI) kijelölésével hamarabb váljon lehetségessé. Az egyszerű rendszerek esetében a felismerés egyszerűen leállítja a további működést, ezzel például elkerülve egy esetleges károkozást, míg bonyolultabb rendszerek esetében mindenkori további cél, a rendszer valamilyen biztonságos és kontrollált, akár működő állapotban való tartása. A veszélyes helyzet folyamatos nyomon követésével, elemzésével megoldható, hogy akár korlátozásokkal ugyan, de egy rendszer továbbra is működőképes maradjon részben, vagy egészben (pl. vasúti rendszerek, repülők berendezései), vagy hozzásegítsük ahhoz, hogy biztonságos működő állapotba juthasson.

A mindenkori cél tehát, hogy az észlelést követően történjen valamilyen erre reagáló tevékenység: valamilyen folyamat elindítása, folytatása, megváltoztatása, vagy leállítása amellet, hogy a hibák, működési feltételek megváltozása jelzésre kerül, vagy további adatokat kér döntési pontoknál.

A rendszercélok elérésénél komoly megszorítást jelent a beágyazott rendszereknél a „speed-power-area” hármas, vagyis a sebesség, az energiaigény és a feldolgozást végző

chip felületméretének egyidejű optimalizálása. Ez olyan megszorítást ad, amelyre érdemes algoritmikus trükköket elővenni, vagy a természetben megtalálható valamely hatékony mechanizmust lemásolni.

1.1 Motiváció

Az objektumszegmentáció a képfeldolgozási lánc egyik fontos építőeleme. Valamely objektum térben való elhelyezkedésének vizsgálata, tulajdonságainak meghatározása, vagy becslése információt ad a további feldolgozási feladatokhoz. Még egyetemi tanulmányaim alatt egy konferencián publikált kutatási feladatomban [4] vizsgáltam, hogy az akkor elérhetővé vált Eye-RIS v1.2 kamerában [5], - mely egy Celluláris Neurális Hálózat (CNN) felépítésű, fókuszsíkbeli analóg szenzor-processzortömb, - milyen hatékonysággal lehet CNN műveletek alkalmazásával megvalósítani előtér-háttér szegmentációt. Ezen kezdeti kutatás jó alap volt a vizuális jelenségek tér-időbeliség vizsgálatának megismerésére és olyan elméleti kérdések feszegetésére, hogy mi értelmezhető előtérnek és mi inkább háttérnek.

De mit is tekinthetünk előtérnek és háttérnek? Mindkettő meghatározása relatív, függ a környezettől, időbeliségtől, kontextustól: ha egy objektum az adott térrészen megjelenik, mint például ahogy leparkol egy autó, akkor mennyi idő az, amikortól háttérnek tekinthető? Ha valami a háttér része, mint például egy hirdetés, akkor annak megváltozása mit jelent? Egy, a városokban előforduló forgó óra az előtér, vagy a háttér része? Vagy egy kitűzött lobogó egy ház oldalában minek számít?

Az előteret talán egyszerűbb lehet megfogni: előtérnek lehet tekinteni minden, a képen megtalálható objektumot, ha az képes elhelyezkedését megváltoztatni, elmozdulni a térben. Egy folyamatosan mozgó tárgy (például egy óra, kitűzött lobogó) inkább a háttérhez tartozó objektumnak számíthat. Általánosságban elmondható, hogy egy felismerési feladat kapcsán aktív cselekvésre jellemzően az előtér objektumainak pozícióváltozásakor lehet számítani. A háttér objektumainak megváltozása is természetesen megfigyelhető, hozzá feladat igazítható.

A látás, és ezen belül a retina működésével kapcsolatban 2008-ban ismertem meg Roska Botond svájci kutatócsoportjának munkáját és eredményeit, melyeket a Nature-ben 2009-ben publikáltak [6]. Kutatásukban az egerekben megtalálható, az objektumok közeledésére érzékeny retinacsatorna sejtjeinek meghatározásával foglalkoztak. A kutatás részleteinek megismerése során az „előtér” viszonyát a „háttérhez” képest át kellett értékelnem,

mivel az előtér itt egy látható és jól felismerhető történés volt, melyben minden más látható gyakorlatilag a háttér részévé vált.

Ezen kutatás izgalmas tématerületet jelentett, így elkezdtem foglalkozni a kutatócsoport eredményeinek modell alapú igazolásával, valamint azzal is, hogy a szimulációs modell szinten túllépve, a rendelkezésemre álló CNN, fókuszsíkbeli szenzor-processzor tömb architektúrán konkrét működő implementációját is elkészítsem, ezzel valós időben is elvégezhető kísérletekre lehetőséget adva. A modellt vizsgálva, valamint a működését megfigyelve kíváncsi voltam arra is, hogy a kapott eredmények a jelenség felismerési képességén túl még hogyan lehetnek hasznosíthatók, milyen más következtetéseket lehet belőlük levonni.

1.2 A disszertáció felépítése

Disszertációmban először áttekintem a vizuális jelenségek érzékelésével és észlelésével kapcsolatos alapvetéseket. Az észlelés kapcsán elengedhetetlen foglalkozni a látással, így áttekintem a gépi, valamint a biológiai látás alapjait. A látó rendszer és ezen belül különösen a retina működése kapcsán kiemelem azokat a legalapvetőbb mechanizmusokat, amelyek ismeretét szükségesnek tartom a retinacsatorna modellben található elvek és történések megértéséhez.

A kutatásom alapjául szolgáló, közeledésre érzékeny egér retinacsatorna, az első emlősökben megtalált erre a jelenségre érzékeny biológiai struktúra. Ennek jelentősége abban áll, hogy egy komplex esemény detekciója valósul meg retina szinten. Az ilyen bonyolultságú eseményeket általában a vizuális kéreg szokta detektálni, de annak késleltetése már elég nagy és ez a történés élettani szempontból fontos annyira, hogy már a retina is képes felismerni, vagyis megvalósuljon a korai látás. Bemutatom, hogy korábban nem emlősök esetén milyen más biológiai alapú modell született a közeledő objektumok észlelése kapcsán. A közvetlenül a retinacsatorna működése alapján készült kutatásom képezi 1. téziscsoportom alapját.

A 4. fejezetben a modell vizsgálata során bemutatom a modellben felfedezett egyik mellékhatást, mely a modell működésitartományát befolyásolta, s egyúttal lehetőséget adott arra, hogy kísérletet tegyek a modell alkalmazásával impulzus üzemben működő fényforrások közül a jelzőlámpa jelzőképének detekciójára. Ezzel kapcsolatos eredményeimet a 2. téziscsoportomba soroltam. Téziseimet az 5. fejezetben rendszerezem és ismertetem.

2 Vizuális jelenségek észlelése

A vizuális jelenségek megfelelő észlelése nagy jelentőséggel bír az autonóm rendszerek döntéselőkészítése során. Az ember a külvilágról információt mintegy 80%-ban látás útján szerzi be. Ez alapján képes előre tervezni, előre látni folyamatokat, reagálni azokra, és az információk összefüggéseinek felfedezésével képes a világot megismerni. [3]

Ebben a fejezetben áttekintést adok az érzékelés és észlelés folyamatáról, kiemelve, hogy hogyan valósul meg ez a látás során.

2.1 Érzékelés és észlelés

Az észlelés egy folyamat, mely az érzékszerveken keresztül érkező információk felvételét (érezékelését), feldolgozását, valamint az ezekre adott választ foglalja magában. [3] Az észlelési folyamatok bonyolultak és sokfélék lehetnek, az „érzet”, vagy „érzéklet” által kiváltott válaszok összetettek. Az utcán sétálva könnyen megragadhatja valami a figyelmünk, például egy ismerős arca a tömegből. Viszont, ha kifejezetten és szisztematikusan keresünk ugyanazt az embert, ugyanabban a tömegben, akkor nem biztos, hogy könnyen rátalálunk. Az érzékleteti mintához is hozzárendelődnek a környezet további ingerei, eseményei, kontextusa, melyből maga az észlelés kialakul.

Az észlelés és érzékelés különbözőségének tárgyalása már 17. századi filozófusoknál megjelent. [7] Az észlelés legfontosabb alapja az érzékelés, alapvetően határozza meg az észlelés minőségét, lehetőségeit. Az érzékelés önmagában is egy összetett és bonyolult folyamat, melynek egyik legfontosabb eleme a receptoroknak nevezett különböző idegsejtek.

Az érzékelés során a fizikai környezet ingerli a biológiai rendszert a különböző receptorokon keresztül. [7] Ennek során például a fizikai nyomás a mechanoreceptorokra hat, míg valamilyen hőhatás a thermoreceptorokra. Az étel szagát, vagy ízérzetét különböző chemo-receptorok fogják fel, míg egy biológiai szervezet a fényhullámokra a fotoreceptorokon keresztül érzékeny. Amikor egy receptort az úgynevezett nyugalmi potenciálban elég erős inger ér (ez az ingerküszöb), akkor ott kialakul az akciós potenciál, mellyel megkezdődik a sejtek közötti kommunikáció ingerületátvitel segítségével, különböző ingerületátvivő anyagokkal. Az egyes receptorok ingerküszöbe egyenként és fajonként eltér egymástól, valamint a sejten megjelenő akcióspotenciál időbeli lefutásában is vannak különbségek. Ez az evolúció folyamata során egyenként kialakult jellemző.

A biológiai rendszerekben a bemeneti ingerek különböző információkat hordoznak, melyek egy része azonnal feldolgozható jelentéssel bír, lényegében véve egy huzalozott utat követ. Az inger-válasz reakciók a különböző biológiai rendszerekben a túléléshez szükséges alapvető viselkedési mintákat tartalmazzák [2]. Ilyenek például a feltétlen reflexek, amelyek esetében az "utasításkészlet" alkalmas arra, hogy az egyedfejlődés során tapasztalt tipikus helyzetekre, vészhelyzetekre valamilyen átlagosan jó, a túléléshez hozzásegítő választ adjon. Maga a reflex ebben a formában nem tekinthető észlelésnek, viszont tudatosulása maga már lehet észlelés, ha megtörténik annak elemzése, hogy milyen eseményre történt válaszadás. Az észlelés lényegi pontja az, amikor realizálódik, hogy egy reakció milyen esemény hatására történt.

A vizuális jelenségek feldolgozására és észlelésére a biológiai szervezetek számos megoldást adnak. Neurális hálózatok oldanak meg olyan alapvető problémákat, mint perspektivikus transzformációk, vagy különféle mozgások észlelése: horizontális és vertikális, növekedés vagy csökkenés, forgás, rövidülés [8]. Ide tartoznak még egyes képjellemzők felismerésére szolgáló megoldások is, amik az élkereséshez, fókuszáláshoz kapcsolódnak.

Hubel és Wiesel a macska látórendszeréhez kapcsolódó agykéreg vizsgálatának eredményeként fogalmazták meg a neuronális szelektivitás elvét. [7] A látókérgi kutatásuk során arra jöttek rá, hogy külön-külön neurális hálózat valósítja meg a retinában a függőleges, vagy éppen a vízszintes vonalak felismerését [3]. A macska V1-ként jelölt agykérgi részében felfedezett mechanizmus rámutatott, hogy az agykéreg idegsejtjeinek egy része az érzéklet meghatározott tulajdonságaira nézve specifikusan lép működésbe. Ezek mind olyan információk, melyekre számos mechanizmus épülhet (objektum felismerése, védekező reakciók, szemlencse homorúságának változtatása, erős fény esetén a szemhéj becsukása, stb.). [7] Ennek kialakulásában és minőségének szempontjából a születési utáni első időszak a legfontosabb [9].

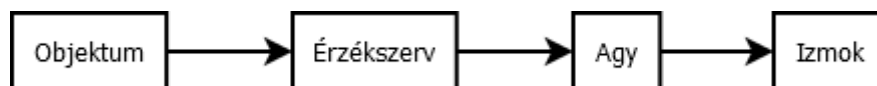
Az agy észlelési jelenségei közül a vaklátás disszertációm fő vonala szempontjából különösen fontos és megalapozza a korai látásra való képességet. A vaklátás a látótérfél valamely részének kiesését jelenti, konkrétan nem funkcionális működését. A vaklátáskor a beteg alapvetően vak, nem képes tárgyakat felismerni, viszont George Riddoch kutatásában azt vette észre, hogy a retinálisan nem sérült emberek esetében, hogy valamely látótér kiesésekor, a látótéren történő különböző irányú mozgások megtörténtét a betegek képesek megmondani [10]. Larry Weiskrantz részletesen vizsgálta is a jelenséget [7], kísérleteit „D.B.” nevű betegével kapcsolatban írta le részletesen [11]. D.B a mozgás irányultsá-

gán túl képes volt egy-egy tárgy további vizuális jellemzőit is észlelni, mint például tárgyakon a kontrasztokat ([11] 12. fejezet). Ez arra utal, hogy ezek kifejezetten fontos jellemzők az észlelés érzet kialakításának szempontjából.

A küszöb alatti észlelés fogalma egy olyan további észlelési jelenség, mely még munkám szempontjából figyelmet kíván. Ide tartoznak azok az észlelések, melyek küszöb alattiak, s melyek jellemzően olyan tárgyakra, jelenségekre jönnek létre, melyről az észlelőnek nincs tudomása, nem tudatos [7]. A fogalom nem jól körül határolt és tudományos oldalról nem jól definiált: elvont szempontból ide szokták érteni az 1-2 filmkockában felvillanó reklámokat, de az illuzionisták is erre építhetik műsorukat, valamint tipikusan ide tartoznak a figyelmet magukra vonó jelenségek.

Az 1. ábra mutatja be az észlelés főbb eseményeit, ahogy a pszichofizika tárgyalja. Ennek kapcsán a fizikai inger és az észlelés között történik kapcsolatteremtés azért, hogy arra valamilyen cselekvési terv készülhessen.

A pszichofizika szerint az észlelés események sorozata, amely az észlelőn kívül eső fizikai világ eseményeinél kezdődik (objektum kapcsán valamilyen jellemző az érzékszervhez jut), lefordítódik mintázattá (érzékszervtől az agyhoz) és a mintázat feldolgozását követően válaszadással fejeződik be (agyból cselekvés/cselekvés abbahagyása az izmok által).



1. ábra Az észlelés eseménysora ([3] 1.1 ábrája alapján)

Az észlelés (vagy más megközelítés alapján az érzet [7]) egyik nagyon fontos sajátossága, hogy alapvetően szubjektív. A feltételes reflexek megléte is ezen alapul, valamint az érzet kialakulása is. Függetlenül a korábbi tapasztalatoktól, tanult reakcióktól, vagyis nem jöhet létre az érzékelt, ismert valóságtól függetlenül. Dr. Kardos Lajos pszichológus idézete ennek kapcsán ide kívánkozik:

„Ha egy ember szemébe bizonyos hullámhosszú fényt sugároznak, azt mondja, hogy kék színt lát. Minden más ember is ezt mondja, ha ugyanolyan hullámhosszú fény érkezik a szemébe. Teljesen ésszerű azt állítani, hogy az adott szituációban mindenki „ugyanazt” látja, „ugyanazt” éli át, látási érzéklete – legalábbis a jegyét illetően – ugyanaz. [...]

Ezt a modern behaviouristák is belátják, de külön hangsúlyozzák: szükséges feltétel, hogy legyenek olyan szituációk, amelyekben mindenki ugyanazt (vagy lényegében ugyanazt) a verbális reakciót adja, ugyanarról az élményről számol be.”[12]

Bármit is érzékelünk, az kontextusba helyezve kap értelmet és ennek megfelelően válhat tudatos észleléssé.

2.2 Látás, gépi látás

A látás összetett folyamat. A biológiai oldalról a látás nem pixel szintű képfeldolgozást jelent az agyban, hanem a retinában helyben „tömörített” információk átadását a központi idegrendszer és az agy számára, melyek ezen érzetekre támaszkodnak a további feldolgozás során, melyből az észlelés folyamata kialakul. A gépi látás olyan eljárások összessége, melynek a végén egy a vizuális információkra támaszkodó rendszer a történésekre képes reagálni. Az ezen eljárásokban alkalmazott megoldások manapság már sok tekintetben a biológiai rendszerekhez hasonló megközelítést alkalmaznak a feladatosztályok megoldására. A következőkben bemutatásra kerül a működésük, valamint a kapcsolódó fontosabb alapvetések.

2.2.1 A látórendszer működése

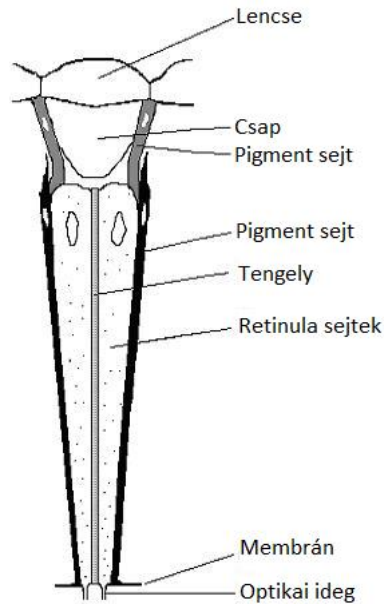
A látószerv felépítése a különböző élőlényeknél jelentős különbségeket mutat. Az egysejtűeknél a sejthártya képes érzékelni a fényt, de a puhatestűek egyes fajainál lehet először konkrét fényérzékelésről beszélni. Az ízeltlábúaknál már százas, vagy akár többes nagyságrendű pontszerű szemből (ún. *ommatidium*, 2. ábra) álló facetta szemről beszélhetünk, míg a gerincesek esetében pedig egy bonyolult felépítésű látórendszerrel találkozhatunk [13, 14, 15, 16].

2.2.1.1 A retina

A látórendszer első és legfontosabb eleme a szem, mely everz, vagy inverz lehet. Ha az érzékelt fény közvetlenül éri a fényre érzékeny sejteket, akkor everz szemről beszélünk, ha pedig valamilyen bonyolultabb struktúrán (pl. lencsén) kell a fénynek keresztülhaladnia, akkor inverz szemről van szó. A szem komoly előfeldolgozási folyamatokat végez az agy és a központi idegrendszer számára, hogy kialakulhasson maga a látásérzet.

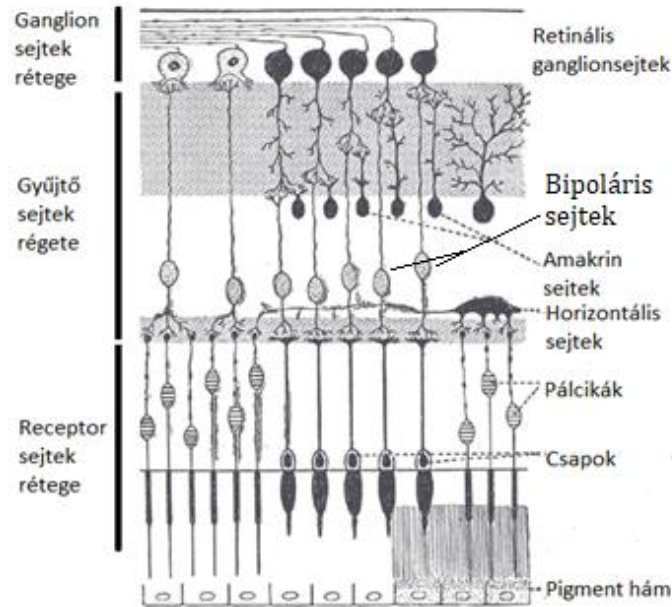
Konkrét látásról a rovarok esetében lehet először beszélni, még ha nem is lehet tudni, hogy agyi szinten mit képesek felfogni a látottakból [17]. A látórendszer a rovaroknál viszonylag egyszerű felépítésű, amelynek összetett szemes változatát a 2. ábra mutatja.

A fényérzékeny sejt az ún. retinula sejt, melyből általában 8 található egy *ommatidium*-ban [17], mint például a sáska látórendszerében [18], míg egy méh esetében pedig akár 9 is lehet [17]. Az *ommatidium* vége közvetlenül a látóidegben végződik, mely az állat központi idegrendszeréhez kapcsolódik [19] [20].



2. ábra Rovar ommatidium felépítése, a retinula sejtet körbe veszik a tengelyt ([19] ábrája és [20] leírása alapján)

A gerincesek esetében inverz szemről beszélhetünk [14] [15] [13]. A gerincesek retinája három rétegű rendszert alkot, melyet a 3. ábra mutat be az emberi retina felépítésén keresztül. Minden gerinces esetében ez a háromrétegű felépítés található meg. Dowling és Boycott részleteiben írja le elektronmikroszkópos [21] kutatásukat a felépítés kapcsán, melyben az egyes rétegeken belül további rétegeket azonosítanak, de az adott rétegben található sejtek feladatai alapvetően hasonlóak, vagy azonosak [14], [15].



3. ábra Emberi retina felépítése ([22] ábrája és [14] leírása alapján)

A szemben a fény az üvegtesten át érkezik a retinára. A beérkező fény energiáját a retina érhártya felőli oldalán megtalálható, fotoreceptorokat tartalmazó réteg alakítja át idegi jelekké. A fotoreceptorok lehetnek különböző hullámhosszokra érzékeny csapok, melyek a színérzékelésért felelnek, vagy a fényérzékelésre alkalmas pálcikák, melyek leginkább a gyenge fényben, szürkületben való látásért felelnek. A csapoknak a színérzékelés kapcsán különböző típusai vannak, melyek különböző hullámhosszú fényekre érzékenyek. Az ember esetében ez három csaptípust jelent, amelyek a piros (~570nm), a zöld (~540nm) és az ultraibolya (~440nm) hullámhosszokra érzékenyek a leginkább. A színek érzékelése, az ún. trivariáns színlátás, az embernél a különböző hullámhosszra érzékeny csapsejtek által érzékelt hullámhosszok szuperpozícióján alapul, valamint legalább két csapsejt által érzékelt hullámhosszokból vett különbség felismerésén. A színélmények már idegrendszer szinten alakulnak ki, az észlelés során ezeket az árnyalat-telítettség-világosság hármasa írja le. [7] [14]

A csapok és pálcikák aránya a különböző fajokban más-más. Retinán való elhelyezkedésükre jellemző, hogy a csapok a retina sárgafoltjának környékén fordulnak elő legnagyobb számban, közepén kizárólag csapok, amely terület az éleslátásért felelnek.

A különböző idegsejtek közötti kommunikáció szinapszisokkal történik, mely során a sejtek között ingerület átvivő anyagok (kémiai hírvivő molekulák) az egyik sejt axonjától a

másik sejt dentritje felé átadásra kerülnek. Ezek a transzmitter rendszerek lehetnek serkentő, vagy gátló hatásúak. A fényérzékelés is ezen rendszer szerint történik.

Sötétben a Ciklikus guanozin-monofoszfát (cGMP) kémiai hírvivő koncentrációja viszonylag magas a receptorsejtben, ekkor a sejten lévő csatornák nyitva vannak, a sejt depolarizált állapotban van. Amikor fény éri a fotoreceptor sejtet, az aktiválja a vizuális pigmenteket (csapnál pszin (protein) és 11-cis retina molekulákat), amik egy enzimet stimulálnak, mely csökkenti a cGMP koncentrációt a citoplazmában, ami bezárja a cGMP-kapuzott csatornákat, amivel megszűnik a sejtnél meglévő Na^+ és Ca^{2+} beáramlása. Ennek hatására a fotoreceptor hiperpolarizált állapotba kerül ($-40\text{mV} \rightarrow -70\text{mV}$), melynek eredménye, hogy a sejt szinapszisában a glutamát neurotranszmitter felszabadulása csökken, ami a retina második rétegében található gyűjtősejteket hozza működésbe. [23]

A gyűjtősejtek alkotta réteget bipoláris sejtek, amakrin sejtek, Müller glia sejtek és horizontális sejtek alkotnak [14] [21]. Ezen sejtípusok közvetlenül kapcsolódnak a fotoreceptor sejtekhez. Szerepük főként a fotoreceptor sejtől érkező jel előfeldolgozásában van, akcióspotenciál kiváltása nélkül. Ezen sejteknek morfológiailag különböző altípusai vannak, melyek jól megkülönböztethetőek egymástól. Ilyen például a bipoláris sejtek ON-OFF működését meghatározó „előjelváltási”, vagy „előjel megőrzési” feladata. Ennek a működésnek az az oka, hogy vannak olyan morfológiai műveletek, melyekhez kivonás jellegű műveletre is szükség lenne, viszont az idegrendszer „elvenni” nem tud, mivel a természetben nincsen kivonás. Oka, az előzőleg említett sejtek közötti kommunikáció szinapszisokkal történő egyirányú megoldása, vagyis ennek megfelelően a „kivonás” is csak ingerület átadással tud megvalósulni a gátló-serkentő mechanizmuson keresztül.

Ha számítástechnikából vett analógiával akarunk élni, akkor ez a működés hasonlít a fixpontos számábrázolás kapcsán ismert, a kettes komplementerképzéssel végrehajtott összeadáshoz, amely során csak összeadást és invertálást alkalmazunk. Ennek a műveletnek a végeredménye pedig a két érték közötti különbség lesz (az első jegy elhagyását követően).

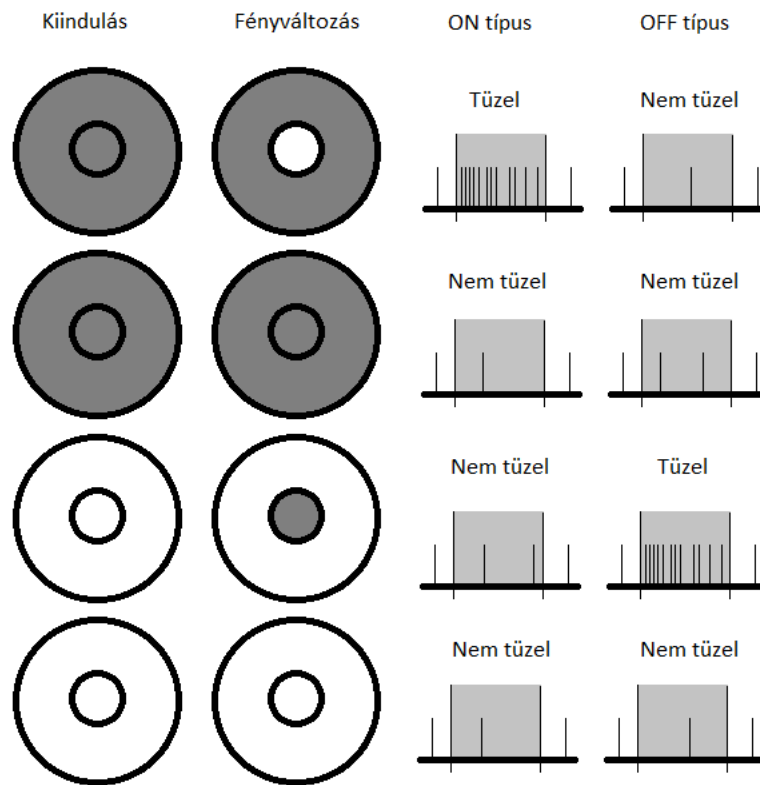
Amikor a retina csapsejtét egy sötétből világosba váltó fényinger éri és azt valamely bipoláris sejt továbbítja, akkor azt ON típusú bipoláris sejt teszi. Ha világosból sötétbe váltó fényinger esetén történik ingerületátvitel, akkor pedig OFF típusú bipoláris sejtről beszélünk.

A retina harmadik rétegében ganglionsejtek találhatóak. Az itt található ún. retinális ganglionsejtek idegnyúlványai alkotják magát a látóideget, és kapcsolódnak az idegrendszerhez. Itt történik meg az ingerületek továbbítása az agy, valamint a központi idegrendszer felé. A retinális ganglionsejtek jelentősége abban áll, hogy ez határozza meg alapvetően magát a látásérzetet [24]. Ezek a sejtek nem csupán a receptorsejtek irányából érkező ingerületgyűjtő funkciót látják el, hanem a látottakról egyfajta előzetes értékelést is adnak az agy és az idegrendszer felé. Ezek a fejezet elején is már említésre került irányultság, perspektivikus érzet, élek felismerése.

Ennek a komoly előfeldolgozásnak az az oka, hogy a fényre érzékeny receptorsejtek olyan nagy információs mennyiséget jelentenének az agy és az idegrendszer számára, hogy egy az egyben való továbbításuk lényegében lehetetlen. A retina így gyakorlatilag nem csak a látott információt „tömöríti”, hanem egyenesen arról ad információt, hogy milyen jellegű történés van a retina egy jól meghatározott régióján, a receptívmezőn.

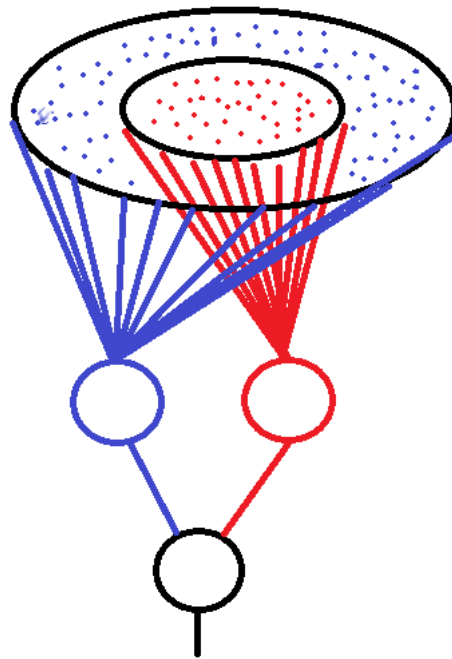
A receptívmező egy olyan térrész a retinán, melyet egy ganglionsejt közrefog és amely területen a gyűjtősejtek egy részével, funkciójának megfelelően kapcsolatban áll, vagyis az adott térrész szelektív látásérzetét adja. A receptívmezők egymással átfedésben vannak, sok, és sokféle ganglionsejt alkotja őket.

A retinális ganglionsejt típusa nagyban meghatározza az átadott ingerület minőségét és átadásának módját. Megkülönböztetünk „ON” típusú és „OFF” típusú válaszra érzékeny ganglionsejteket [14] [3], attól függően, hogy a „sötétből világosra”, vagy a „világosból sötétre” reagálnak ingerületátvitellel (4. ábra). Ez lényegét tekintve a kontrasztok változásaira való reagálás képessége, ami a tárgyak érzékelésének, azok határaik megtalálásának egy fontos érzékelő eleme. Az inger-válasz kapcsán fontos megjegyezni, hogy az ingerületátvitel egy burst jellegű tüzelés, amely a változás minőségének megfelelően alakul, az időben a változástól számítva lecsengő. Ennek egyik oka az ingerületátvivő anyagokkal való telítődés. További tulajdonsága, hogy nyugalmi állapotban is van egy állandónak tekinthető időnkénti aktivitásuk. Ennek forrása a retinális interneuron hálózat.



4. ábra Fényváltozás időbeli hatása ON center és OFF surround ganglionsejtek esetén. Nem tüzelésnél időnkénti aktivitás van, tüzelésnél burst jellegű, csökkenő aktivitás.

Felépítésükre jellemző, hogy alakjuk körhöz hasonlítható, valamint az is, hogy a sejt által lefedett receptívmező belseje ON típusú, míg periferiája OFF típusú fényváltozásokra reagál aktivitás-mintázattal, vagy fordítva (5. ábra). Ezt a koncentrikus elrendezést hívjuk a sejtmező antagonista felépítésének. A ganglionsejtek, mivel ingerület gyűjtő funkciót látnak el, ezért nyugalmi állapotban is mutatnak a típusra jellemző aktivitást, függetlenül attól, hogy megvilágított állapotban, vagy sötét állapotban vannak. Ez a jelenség az, mely során az ember a teljesen sötét helyen is képes „látni valamit”. Ekkor ugyanis minden sejt aktivitása lecsökken és az egymáshoz képesti relatív viszony miatt kelt látásérzetet az idegi aktivitás.



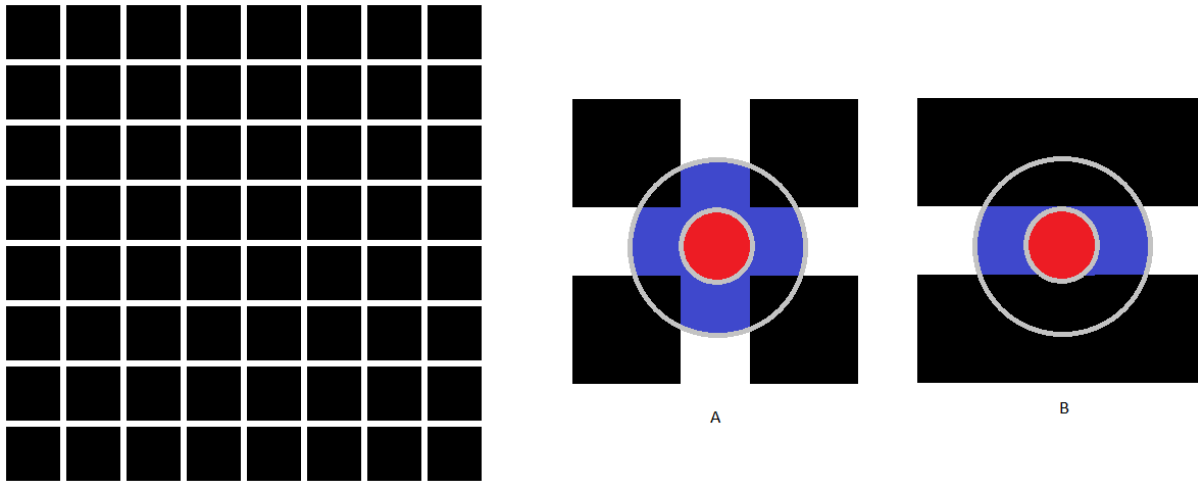
5. ábra Antagonista felépítés, a piros és a kék ellentétes (ON, vagy OFF) információt gyűjt

A neurobiológiában a serkentő-gátló mechanizmusok kapcsán a leglátványosabb és legmegfoghatóbb a retina működése, ugyanis látványos, optikai csalódást tud okozni a receptív mező ON-OFF működése, ha az antagonista felépítést is figyelembe vesszük. Ennek egyik legáltalánosabb példáját a Hermann-rácscsal lehet bemutatni (6. ábra). A fehér alapon rácspan elhelyezkedő fekete négyzetek sarkainál feltűnik, hogy sötét árnyalat látható azokon a helyeken, melyek nem az ún. fikszióspont, vagy környékére esnek.

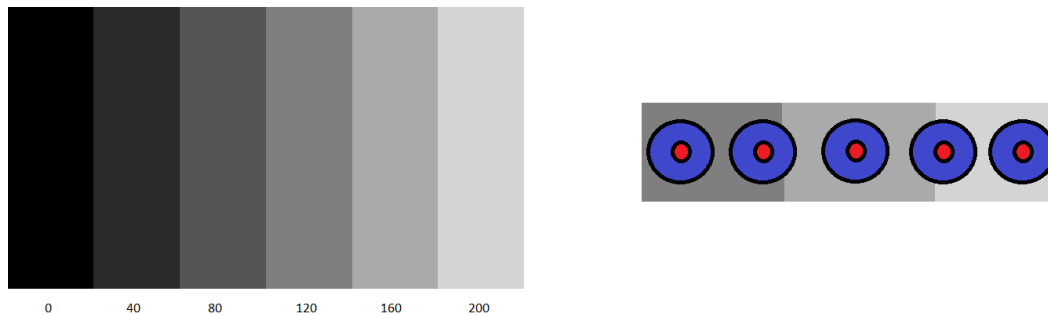
A jelenség egyik általános magyarázata, a látótérben a ganglionsejtek által alkotott receptívmezők hol helyezkednek el. A retina közepén kisebb tartományt ölelnek fel, míg a periféria felé haladva egyre nagyobbakat ([3] 91-92.), itt a receptorok száma is jelentősen csökken már, vagyis az észlelést meghatározza az is, hogy a retinasejtek mekkora halmozattól fogadnak információt.

A rácspontoknál a receptívmezőre több fény esik az OFF választ adó részekre, mint a négyzetes területet fedő részeknél így a beérkező jelek összege a ganglionsejten belül kisebb lesz. A rácspontokban az ON-ra érkező válasz a vonalak mérete miatt nagyjából ugyanakkora, mint az oldalak mentén, viszont a sötét ablakok miatt a ganglionsejt szélén fokozottabb gátlás éri. Ennek eredményeként jön létre ezeken a helyeken ez a lényegében nagyobb válasz, mely az agy számára úgy jelenik meg, mintha a rácspontokhoz képest az

oldalak mentén világosabb lenne a vonal, vagyis a látórendszer relativizál [3] [7] (6. ábra A és B).



6. ábra A Hermann rácsra tekintve a rácspontokon a nem fókuszban, vagy közelében lévő rácshelyeken sötét pontok láthatóak. ([3] alapján).



7. ábra Mach sávok, az egymás mellé rakott különböző világosságú mezők, a mezőnként állandó világosság ellenére átmenetesnek tűnnek. ([3] alapján).

A fényerősség érzetének a különbségét szemlélteti a Mach-sávot bemutató 7. ábra, melyen az egyes homogén világosságú oszlopokat átmenetesként lehet érzékelni, az egyes sávok szélei csíkozottnak hatnak, főként a sötétebbek esetén. Ez szintén a receptívmező antagonizmus eredményére vezethető vissza. A retinán megjelenő fény túllép az egyes retinális ganglionsejtek mezején. A 7. ábra jobb oldalán szereplő receptívmezőket feltételezve a középső és a két szélső, homogén részre eső ganglionsejtek válasza a reprezentált világosságnak megfelelően alakul. A maradék két esetben viszont a receptívmező ON közepe jól meghatározottan beleesik valamely sávba, míg az OFF szélének jelentős részére egy másik sáv fog hatni. Emiatt az azonos térrészben lévő teljes receptívmezőhöz képest az adott

ganglionsejt válasza más lesz az adott térész vonatkozásában, nagyobb vagy kisebb. Ennek eredménye, hogy sötétebbnek vagy épp világosabbnak tűnik az adott terület a másik oldali térrészhez képest. Ha az ábrát felnagyítva tekintjük meg, akkor ez a gradiens jellegű illúziós jelenség eltűnik, mivel a receptívmezők nagy része már látszólag csak egy sávot észlel.

A kiterjedésbeli különbség kapcsán még meg kell említeni, hogy léteznek nagysejtes és kissejtes ganglionsejtek is. A nagysejtes ganglionsejtek alkotják jellemzően a perifériás nagyobb receptívmezőket, míg a kissejtesek pedig a *fovea* környékén jellemzőbbek, kiterjedésük is kisebb. A nagysejtes ganglionsejtek idegi axonja vastagabb, és ennek okán gyorsabban jut el az idegi impulzus az agyhoz. A kisebb tárgyakra, vagy a fixációs pontokra (pl. a Hermann rácson egy megfigyelt rácspontra, ahol nem látható a szürkesség) a kissejtes ganglionsejt érzékenyebb, míg a nagysejtes a kontrasztkülönbségek viszonyaira érzékenyebb. [3]

A ganglionsejtek hálózata nem érzékeny a homogenitásra [14], így a valóban érdekes régiók mentén állhat elő a világról alkotott előzetes információhalmaz az agy és a központi idegrendszer számára.

2.2.1.2 A látópályák és feldolgozásuk

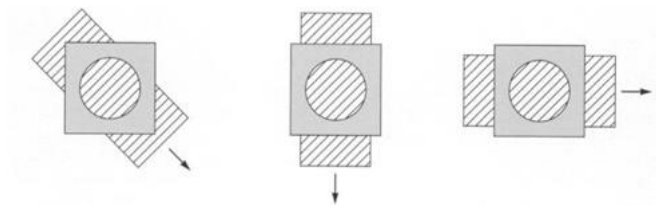
Az emlős retina ganglion sejtjeinek axonjai alkotják a látóideget, ezek a közvetlen bemenetei az agyi feldolgozásnak. A látóidegen belül az egyes rostok rendezetten gyűlnek össze és futnak az agytörzs felé. A rostok fele a szemmel azonos agyféltekére fut be, míg jelen másik fele az agy másik féltekére kereszteződik. Az axonok nagyjából 80%-a az oldalsó geniculatus mag (röviden LG) néven ismert sejtcsoportnál végződik, a többi pedig a közepagy más struktúráiban. [3]

Az LG réteges szerkezetű struktúrában helyezkedik el a két agyféltekén és a retina egy reprezentációját hordozza. Két nagysejtes (magnocelluláris) és négy kissejtes (parvocelluláris) réteg alkotja az embernél. A sejtek mérete az érzékenységet határozza meg. A mozgás sebességre például a magnocelluláris réteg érzékeny, míg a látásélességnél a parvocelluláris réteg működése a meghatározó. [3]

A retina szomszédos területeiről származó információk az LG-ben is egymással szomszédos sejtekhez kapcsolódnak. Emiatt ezt a struktúrát retinotopikus térképnek is nevezik. Az LG sejtjei is kör alakú receptívmezőkbe szervezettek, hasonló antagonisztikus kölcsönhatásokkal, viszont a periférián a gátlás itt sokkal nagyobb. Az LG-ből az információk a

látókéreg felé mennek tovább, ahonnan feldolgozást követően visszacsatolásként visszatérnek. A látásérzet kialakulása alacsonyabb rendű fajoknál az agytörzsben történik, míg magasabbaknál inkább az agykéreg felel érte. [3] [25]

Az elsődleges látókéreg a Brodman 17, vagy V1-ként is ismert agyterület, melyek mellett további V# jelű „Vizuális” területek vannak. Ezek a kérgi területeken lévő sejtek felelősek meghatározni például olyan látható jellemzőket, mint az irányultság (V1 régió), binokularitás (V1 régió), színek (V1 és V4 régiók). Az agykéregben emellett a pravocelluláris és magnocelluláris pályák alkalmasak számos specializált észlelésre, úgymint a textúra-észlelésre, mintázatészlelésre és élességgel, kontrasztokkal vagy éppen a villogások észlelésére (V1-V4 régiók vegyesen). A V1-V2 régiók felelősek az illuzórikus kontrasztok felismeréséért, melyek további agykérgi területekre futhatnak tovább. A V5 régió például a mozgásérzékeléssel hozható főként kapcsolatba, valamint ennek kapcsán az apertúra probléma megoldása is itt történik [25]. Az apertúra probléma a receptívmezők kapcsán egy fontos jelenség, mely az észlelést befolyásolja. Önmagában a receptívmezőből nem megállapítható a környezet ismeret nélkül, hogy milyen irányultságú mozgás történik: álló háttér látszik-e, vagy mozgó előtér. Nem egyértelmű ekkor az sem, hogy egy mozgás valós iránya és a látszólagos iránya megegyezik-e (8. ábra).



8. ábra Apertúra probléma vizualizációja [26]

Szintén ezen V5 régió dolgozza fel a mozgás parallaxis jelenségét: ez a tárgyak egymáshoz viszonyított helyzetének változásainak feldolgozása. A fejünk forgatásával a tárgyak relatív helyzetét nem biztos, hogy meg tudjuk mondani, míg előre-hátra, vagy oldal irányban elmozdulva képesek vagyunk megmondani: a gyorsabban mozduló tárgyakat közelebbinek, a lassabbakat távolabbinak érzékelteti egy az agyi terület, az elmozdulás alapján relativizál. [25] Mindezen felismerési folyamatok összességéből áll elő a látás érzet és észlelés.

2.2.2 A gépi látás folyamata

A magyar nyelvben a „gépi látás” fogalom nem azonos az angol „Machine Vision”-nel, valamint nem azonos a „Computer Vision”-nel sem, ugyanis az elnevezés mind a Machine Vision, mind a Computer Vision témakört összességében takarja, mint „*általános gyűjtőfogalom*” [27] [28].

Angol nyelvterületen van konkrét különbség a kettő között, bár, mint ahogy az a Global Association for Vision Information szervezet tisztázó cikkéből kiderül, ott sem egyértelmű a használata [29]:

„Often thought to be one in the same, computer vision and machine vision are different terms for overlapping technologies.”

A két fogalom által lefedett módszerek és technológiák valóban rendelkeznek közös részterületekkel, mint például a hardvereszközök, képalkotó technológiák és a képfeldolgozási algoritmusok területe.

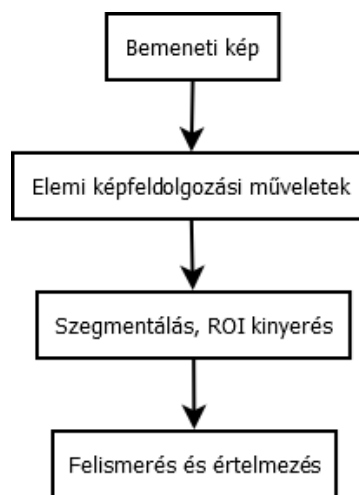
A Machine Vision [30] azon képfeldolgozási megoldásokat takarja, melyek leginkább az ipari technológiákban használatos és az „emberi látáson” túlmutató, tömegfeldolgozás jellegű feladatokra kívánnak algoritmikus megoldást adni.

A Machine Vision témakörébe tartozó feladatosztályok az olyan egyszerűbb feladatoktól indulnak, mint a kódok felismerése és ezek osztályozása (például postai irányítószámok, vonalkódok tömeges feldolgozása). Valamivel szofisztikáltabb feladatokat jelentenek az optikai alapon pozícionáló berendezések, mint amilyenek a palackozóüzemekben található címkéző berendezések, melyek a palackok optikai jelölése (pl. domborulatok az üvegen) alapján képesek felragasztani címkét a megfelelő helyre. A vizuális minőségfelügyelet kiértékelése kapcsán készített berendezésekben pedig sok esetben szükséges bonyolult algoritmusok implementálása (pl. hegesztési varrat ellenőrzése, gyógyszergyárban a tablettákon lévő jelölés ellenőrzése).

Szintén a Machine Vision témakörébe illeszkednek azok a feladatok, melyek emberi döntések elősegítéséhez adnak információkat. Ilyenek a modellező alkalmazások, valamint a diagnosztikai berendezések. Ez utóbbiak egészen komoly elvárásokat támaszthatnak például orvosi területek képalkotó eljárásokat alkalmazó diagnosztikai területén. Egy ultrahangos berendezés a születendő gyermek testrészeinek méreteit képes kijelölni és felmérni, ez alapján pedig megbecsülni súlyát, fejlettségének állapotát.

A Computer Vision témakörbe ezzel szemben sokkal inkább a képen látható objektumok, jelenségek számítógépes értelmezése tartozik. Ide sorolhatók tipikusan a mintafelismerési, objektumfelismerési és azonosítási feladatok térben és időben, melyekhez valamilyen döntési feladat is társul. Egészen nehéz feladatosztály a történések vagy jelenségek észlelése és megfelelő értelmezése egy képfolyamon. Az értelmezés már túlmutat a Computer Vision témakörén, viszont nagyban támaszkodik a kinyerhető adatokra, így a mesterséges intelligencia témaköre szorosan összefügg a Computer Vision területével.

A gépi látás Machine Vision és Computer Vision oldalról is jól definiált folyamatok együttesét jelenti. Mindkét terület a digitális képfeldolgozási műveletekre épül. Alapvetés, hogy bármely képfeldolgozási művelet alkalmazása egy képen, a képi információhoz nem tud hozzáadni, csak megtarthatja az információtartalmat (pl. kép tükrözése), vagy elvehet belőle (pl. zajszűrés). Szerepe így a ráépülő folyamatok számára való adatelőkészítés. A folyamat leírása különböző részletességgel kerül ismertetésre a képfeldolgozást feldolgozó irodalmában, attól függően, hogy egy-egy feladatot mennyi és milyen alfeladatokra bont le ([30], [31], [32]), milyen megközelítést alkalmaz. Ugyanakkor három jól meghatározható lépése van, melyek egymás után hajthatók végre (9. ábra), ezt a folyamatot lehet általános képfeldolgozási láncnak hívni.



9. ábra Általános képfeldolgozási lánc

Az alacsony szintű eljárások [31], vagy előfeldolgozó eljárások [30], olyan alapvető műveleteket jelentenek, melyek közvetlenül a képkészítés módszerével függenek össze. A képkészítés lehet digitális, vagy analóg (ekkor digitalizálni kell). Lehet két és háromdimen-

ziós képkalkotó eljárás használni, mozgóképet, vagy állóképet felhasználni, a feladat meghatározza az alkalmazható eljárásokat. A feladattól függ, hogy milyen képkalkotóeljárással történik a felvétel készítése, mire kell felkészíteni. A képkészítést követően van helye a legegyszerűbb elemi képfeldolgozási műveleteknek, melyek az előfeldolgozási lépés részei („preprocessing”), így például a szintér váltásoknak, képjavításnak, sajátosság kiemelésnek, esetleg egyszerű morfológiai műveleteknek.

A középső szintű feldolgozási eljárások már egyszerűbb tulajdonságkinyeréssel foglalkoznak: éldetektáló, képfokozó, képet helyreállító eljárások, képtömörítés, de itt jelennek meg a képleírók a kép egészére, mint a Fourier transzformáció, vagy egyes területekre, mint a textúra felismerés és alakzat felismerés, vagy leírók meghatározása, mint például a HOG, vagy a SIFT. Az itt megszülető eredmények kerülnek a magasabb szintű feldolgozási eljárásokhoz, mint képleírók.

A magas szintű feldolgozási eljárásoknál beszélhetünk a ROI-ban található elemek osztályozási feladatáról, melyekhez főként a mesterséges intelligencia témaköréből vett eljárásokat lehet használni: neurális hálós osztályozók, döntési fák, statisztikus tanuláselmélet kernelgépek, az SVM-ek, AdaBoost, melyek akár nemlineáris problémák megoldására is képesek. Az osztályozási feladat során kerülhet a kép egy-egy részéhez valamilyen leíró jellemző, mely aztán további döntési folyamatok bemenete lehet és ezen a szinten kerül elő a szemantikai osztályozás is.

Mozgóképek esetén a változás felismerése, ezen belül is a mozgás felismerése összetettebb probléma, mivel számolni kell a felvételen lévő objektumok elmozdulásával, megváltozásával (mozgás és változás detektálás), vagy a kamera elmozdulásával, vagyis egy vizuális elmozdulással (optikai áramlás felismerése). Az optikai áramlás az a jelenség, mely során egy objektum látszólag elmozdul, viszont ez nem azonos azzal, mikor egy objektum ténylegesen elmozdul (valós elmozdulás). Az optikai áramláshoz pedig a biológiai rendszerek-nél bemutatott appertúra probléma is kapcsolódik. [33]

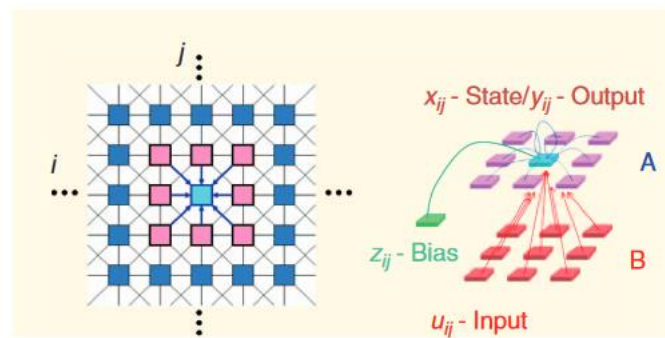
Ezen a ponton válik el érdemben ketté a Machine Vision a Computer Visiontól. A Machine Vision esetében a ROI algoritmikus gyors kiértékelése történik meg, lényegében valamilyen tény felismerése és értelmezése, melyre beavatkozás építhető. A Computer Vision esetében viszont az értékelést további értékelés követheti, más jellemzőkkel együttesen.

2.2.2.1 Látás inspirálta megközelítések

Az elmúlt évtizedekben a látórendszer működésének inspirációja több összetett képfeldolgozási architektúrális modellre hatott, így a neurális hálózatok egyszerűsítésekként 1988-

ban leírt Celluláris neurális hálózatokra (CNN) és a néhány évvel később leírt Konvolúciós neurális hálózatokra (ConvNet).

A CNN elrendezés Leon Chua professzor és Liu Yang neveihez fűződik, amely egy párhuzamos számítási megoldás. Az 1988-ban leírt elv [34] a neurális hálózatok egy egyszerűsített felépítésű megoldását takarja. A Hopfield által 1982-ben leírt neurális hálózattal [35] szemben, itt egy néhány réteges, jellemzően mátrixos elrendezésű hálózatról beszélünk, melynél az egyes csomópontokban lévő elemek csak a szomszédos, vagy nagyon közeli feldolgozási elemekkel állnak kapcsolatban, ahogy a mutatja egy konkrét megvalósulás esetén.

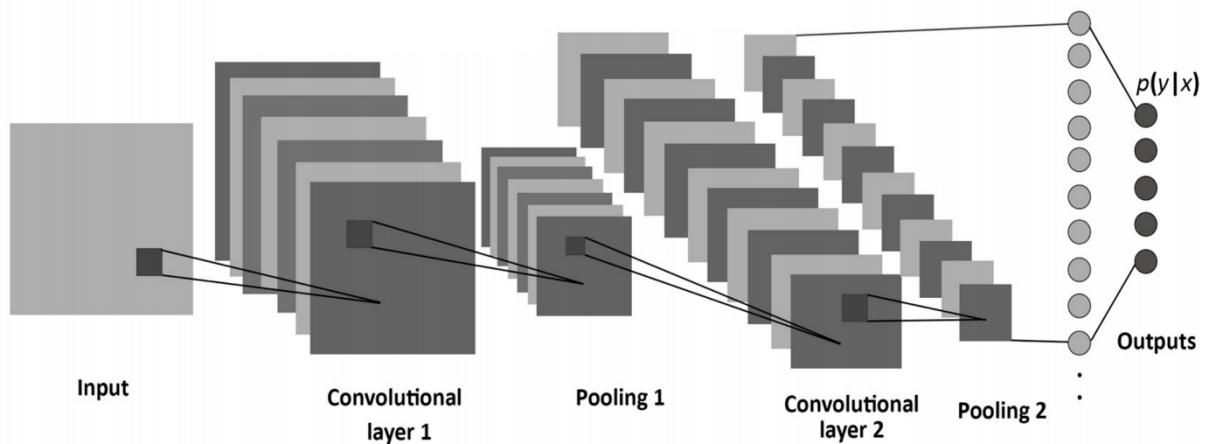


$$\dot{x}_{ij}(t) = -\frac{1}{\tau}x_{ij}(t) + \sum_{k,l \in S_r(i,j)} \mathbf{A}_{ij,kl}y_{kl}(t) + \sum_{k,l \in S_r(i,j)} \mathbf{B}_{ij,kl}u_{kl}(t) + z_{ij}$$

10. ábra CNN mátrixos elrendezése és az állapot egyenlet ([36]ból átvéve)

Egy csomópont, amely egy képpontnak feleltethető meg (10. ábra), melynek értéke a CNN állapot egyenlettel számolható. Az ún. A és B template mátrixok képviselik a CNN műveleteket. Ezek számos alapvető képfeldolgozási műveletre léteznek, így a legtöbb alapvető képfeldolgozási művelethez létezik template (élkeresés, zajszűrés, stb.). A CNN elvet felhasználva készült már retinamodell [37] [38], így a modellezés kapcsán bevált architektúráis felépítésről van szó.

Az utóbbi évtizedben széleskörben elterjedt a Konvolúciós neurális hálózat (Convolution Neural Network, a továbbiakban ConvNet, de mára a CNN elnevezés inkább erre vált elterjedté) alapú algoritmusok használata.



11. ábra ConvNet elvi működése [39]

A ConvNet a mély tanuláshoz kapcsolódó neurális hálózatok közé tartozik, képfeldolgozásban terjedt el alkalmazása, mely hasonlít a látórendszer működésére. A ConvNet kihasználja a felvételek több dimenzióját (pl. RGB színcsatornák), melyeken egy konvolúciós lépést hajt végre egy a feladatra meghatározott konvolúciós kernellel. Ez a lépés lényegében az előfeldolgozási fázis zajszűrésének is megfeleltethető. Ezt követően egy tömörítési lépés következik az ún. pooling réteggel, mely egy maximum keresést jelent a konvolúciós lépés eredményén, majd ezt kisebb térre képezi le. Ezzel a pooling réteg egy absztrakt rétegekként jelenik meg. Egy vagy több konvolúciós-pooling lépés követheti egymást, mely eredménye végül valamilyen döntési, osztályozó algoritmusnak adhatók át, így teljesen csatolt neurális hálózatnak, vagy Szupport Vektor Gépnek. Ez a működés nagyon hasonlít az előző fejezetben bemutatott biológiai látórendszer működésével, ahol hasonló tömörítési lépések történnek (ganglion sejtek a konvolúciós lépést hajtják végre), hasonlóan párhuzamos feladatokkal (különböző célú ganglionsejtek és agykérgi területekre való érkezés, ami a pooling lépésnek feleltethető meg). [32]

A ConvNet elmúlt évtizedes előretörésében nagy szerepet játszott a feldolgozáshoz nélkülözhetetlen elérhető számítási teljesítmény az újabb generációs GPU-kban. A ConvNet alkalmazása számos felismerési feladatra ad hatékony megoldást, jelenleg a vezetést támogató rendszerek is mind inkább építenek ezen módszerre más mély tanulások mellett.

A CNN sokkal inkább takar egy általános architektúrális felépítést, mint a ConvNet. A CNN mivel Turing-teljes, így a ConvNet megoldására képes [40]. A ConvNet konvolúciós rétegének számolása egyértelműen megoldható, míg a második pooling réteg számolása

a CNN GLOBMAX templatejével, valamint a teljesen csatolt réteg eredményének számolása skaláris szorzással lehetséges [40].

A gépi látást párhuzamba állítva a biológiai látórendszerrel, a látórendszerben az alacsony szintű képfeldolgozási műveleteket a receptorsejtek (felvétel készítés) és a gyűjtősejtek (előfeldolgozás) végzik el. A középső szinten elhelyezkedő műveletek alapvetően a ganglionsejt rétegben történnek meg, vagyis itt történnek a különböző tulajdonságkinyerések [41] [42]. A magas szintű feldolgozást már az idegrendszer és az agy végzi el, ezek adják a világról a legösszetettebb észlelést.

2.3 Gyors képfeldolgozás igénye beágyazott rendszerekben

A biológiai látórendszer felépítése párhuzamos ingerületátadásokkal operál. Számos, a látórendszerhez köthető vizuális jelenségre a biológiai szervezet azonnali választ képes adni veszélyek, veszélyhelyzetek elkerülésére. Felmerül a kérdés, hogy a tipikus képfeldolgozási láncban történő feladat megoldásához képest, a természetben is előforduló alapvető túlélési problémákra, a látórendszerben milyen megoldás található? Ezek mennyire hatékonyak, és milyen relációban állnak a szokásos képfeldolgozási láncsal? Ilyen tipikus veszélyt jelentő helyzet a közeledő objektumok érzékelése, mely a menekülésben és a navigációban egyaránt fontos szerepet játszik.

A legszámításintenzívebb feladatoknak a magasszintű, bonyolult képfeldolgozási feladatok megoldása számítanak, amit 2006 környékéig általános célú processzorokkal valós időben nem lehetett elvégezni, mivel a képfeldolgozási műveletek szinte kivétel nélkül párhuzamos műveletvégzést igényelnek. Ekkor jelentek meg általánosságban az első többmagos, párhuzamos feladat végrehajtásra alkalmas GPGPU-k (NVIDIA GeForce 8 sorozat), megadva a valódi párhuzamos számítási algoritmusok implementálásának lehetőségét, akár komolyabb algoritmus könyvtárak létrehozásával (első elterjedtként 2006-ban a CUDA [43], majd a nyílt forráskódú OpenCV). Az elmúlt években jelentek meg azok a GPGPU chipek, melyek alkalmassá váltak beágyazott rendszerekbe való alkalmazásra. Egyik úttörőjük az Nvidia volt a Jetson családjával 2014-ben.

Az Nvidiát megelőzte autóiipari célirányos fejlesztések kapcsán a Mobileye, mely ezen beágyazott számításiigény kielégítésére, kritikus számításaikhoz saját chip kidolgozásával kezdett foglalkozni már a 2000-es években, melynek fontosabb elemeiről a 2.3.2. fejezetben lesz szó.

Szinte az NVIDIA GeForce8-al egy időben jelent meg az egyik legösszetettebb CNN szenzor-processzor tömb, a Q-Eye chipet tartalmazó Eye-RIS [5], mely már alkalmas volt beágyazott rendszerekre épülő fejlesztéséhez. A CNN alapú megoldásoknak korábban már voltak hírnökei (pl. ACE4k [44], vagy az ACE16k [45], Xenon [46] chipek), de ezek csak kevés célfeladatra voltak jól felhasználhatók. A Q-Eye chip köré épített kompakt rendszer viszont (Altera NIOS II FPGA-val) valóban jól használható volt CNN alapú algoritmusok készítésére.

Az Eye-RIS-hoz hasonló kisméretű, fókuszsíkbeli processzoroknál lehetőség van in-tenzív párhuzamos számítást igénylő képfeldolgozási feladatok valós idejű végrehajtására is. Ez megteremti annak lehetőségét, hogy hatékony algoritmikus választ lehessen adni

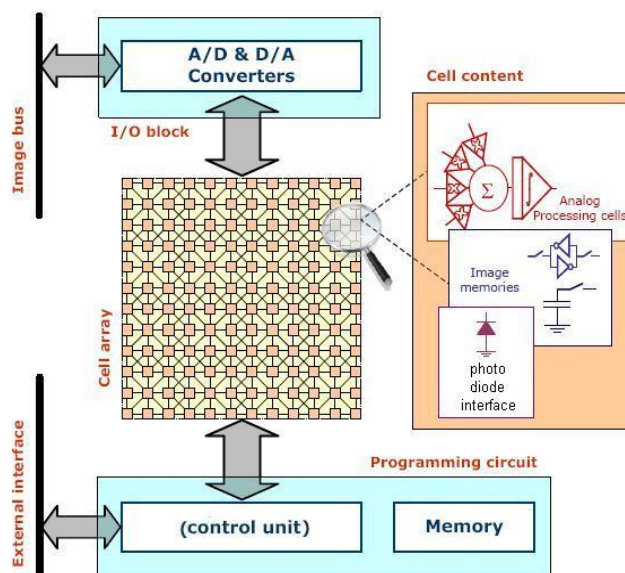
alapvető számítógépes látással kapcsolatos problémák részben, vagy egészben történő megoldására beágyazott rendszereken is – tipikusan a manapság fejlesztésintenzív autóiipari alkalmazásokban jelent ez lehetőséget.

2.3.1 Eye-RIS szenzor-processzor tömb

Az Eye-RIS rendszert a spanyol AnaFocus cég fejlesztette, mely egy általános célú processzorból (az Altera cég NIOS II-es FPGA-ra épülve) és az analóg Q-Eye chipből áll, mely egy fókuszsíkbeli szenzor-processzor tömb (FPSP), CNN elrendezésben.

A Q-Eye chip, egy CNN-re épülő CNN univerzális gép (CNN-UM), feldolgozó egységként egy-egy képérzékelő szenzorral, melynek alapjait [47] [48] Roska Tamás és Leon Chua professzorok 1992-1993 között dolgozták ki. A CNN univerzális gépek analóg processzortömbök első chip implementációjával a Berkeley egyetem foglalkozott még 1994-ben (12×12), nagyobb méretű implementációjával Angel Rodriguez Vázquez sevillai kutatócsoportja foglalkozott az egyik legintenzívebben, mely az első nagyobb felbontású (128×128) megoldással 2003-ban jelent meg, melyet a sevillai AnaFocus gyártott. Ennek következő, már QCIF képfelbontású továbbfejlesztett verziója, a Q-Eye chip 2008-ban jelent meg.

A processzor architektúrális felépítését mutatja be a 12. ábra. Az általános célú processzor a kommunikációért és a vezérlésért felel, míg a Q-Eye chip egy 176×144-es, jelfeldolgozó processzortömb. A Q-Eye minden processzoregységéhez egy képérzékelő szenzor is tartozik (egy pixel), valamint a processzoregységek a szomszédosokhoz kapcsolódnak, 8 szomszédsággal, 3×3-as CNN jellegűen.



12. ábra Az Eye-RIS rendszer felépítése, a Control Unit tartalmazza az Altera Nios II FPGA-t [5]

A Q-Eye chipben 8 bites analóg processzorok vannak. Egy utasítás egyszerre több adaton hajt végre, így SIMD elvű processzor. Minden processzoregységnek van analóg memóriája (LAM) és digitális memóriája (LLU). Az analóg memóriához a tartalmát frissítő áramkör nem áll rendelkezésre, így tartalma az időben változik, „romlik”, mellyel a feladatok elvégzése során számolni kell. Az analóg memória főként képek tárolására alkalmas, míg a logikai memóriák a különböző morfológiai műveletek bitmaszkjainak tárolására alkalmasak.

A processzoregység utasításkészlete tartalmaz logikai műveleteket, alapvető I/O műveleteket, de jellegéből adódóan főként képfeldolgozási műveletek végzésére készítették fel. A rendszerhez szállított programozási nyelv és környezetben a Q-Eye chip az FPP nevű funkcionális programozásnyelvvvel programozható, melynek vezérlésére ANSI C-ben írt nyelv ad lehetőséget az általános célú processzoron.

A rendszer 10 000 fps feletti képfelvételi sebességre képes, valamint feladat függvényében akár 1 000-10 000 fps sebesség mellett feldolgozási és döntési feladatokra, mindezt 100 mW fogyasztás mellett.

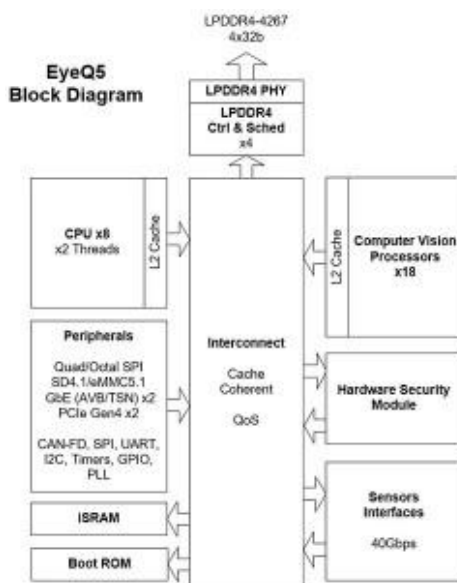
2.3.2 Elterjedt megoldások napjainkban

A Q-Eye chip az Eye-RIS v1.3 megjelenése óta a chip 2014-ben a Toshiba az SPS02 típusú kamerájába is bekerült, de egyelőre továbbfejlesztett változata a v2 nem jelent meg és jelenleg is ez tekinthető a legfejlettebb CNN-UM megoldásnak. Sajnos általánosságban véve problémát jelent a CNN megoldásokban az analóg zaj. Az utóbbi pár évben a CNN

spintronikai megközelítésével is elkezdtek foglalkozni, de jelenleg ez az irány még kutatási fázisban van [49] [50].

Manapság az egyik legfontosabb téma, mely az iparági fejlesztéseket hajtja, az autóiipari alkalmazás, melyben az ADAS megoldásokon túl, az önvezető autó megvalósítása az általános gyártói cél. Ennek megfelelően a legnagyobb beszállítók igénylik a hatékony, beágyazott rendszerekben alkalmazható megoldásokat és itt érhető leginkább tetten ezen beágyazott rendszerek általános elterjedésének lehetősége.

A System-on-Chip (SoC) rendszerek tekintetében integrált megoldások jelentek meg a 2010-es évek óta, melyek mind inkább a mély tanuló algoritmusok (Deep Learning Algorithm - DLA) irányából közelítik meg a különböző problémákat. Az olyan beágyazott, biztonságkritikus rendszerek esetében, melyek ma az autóiipar számára készülnek, így például az önvezető autók számára a Mobileye EyeQ4 megoldása, már 3W mellett 2,5 teraflop számításiteljesítményre képes, mély tanuló algoritmusokat és grafikai modelleket használ fel, melyek egyszerre 8 kamerából másodpercenként 36 képkockát is képesek akár feldolgozni. A chipjük számos hardveres gyorsítási megoldást támogat, az adatlap a DLA-t hardveresen támogatja. [51] Kutatásom korai szakaszában még az EyeQ2 platformra épülő gyalogosfelismerési piacon megjelent megoldásukat gépkocsiban ülve kipróbáltam, ez került bele a Volvo S60 gépkocsikba is radarral kombinált megoldásként. Működése már akkori állapotában is előremutató volt.

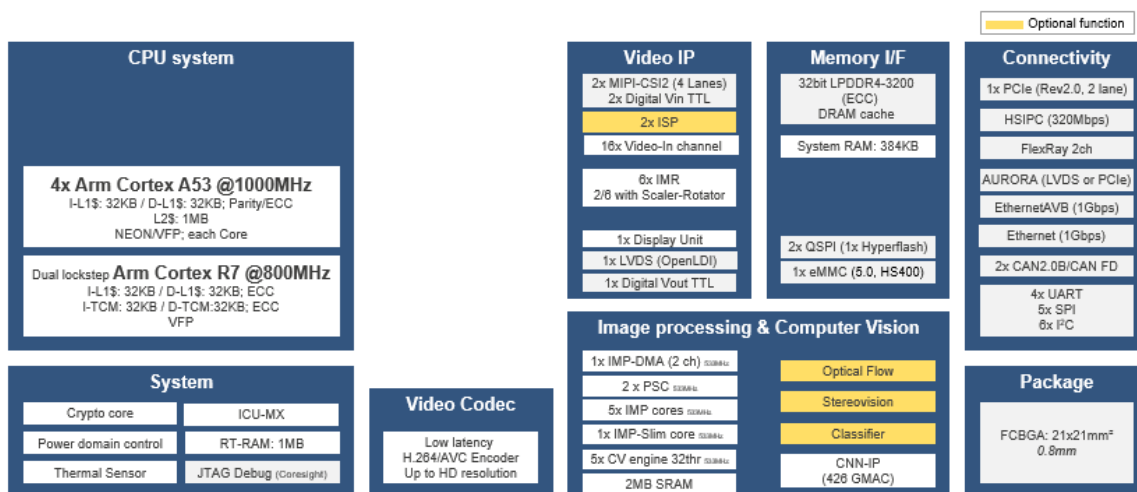


13. ábra Az EyeQ5 tervezett architektúrája [52]

A Mobileye-ban alkalmazott megoldások az emberi látás alapján készülnek, a chip a mély tanuló algoritmusokat is támogatja. A céget 2017-ben felvásárolta az Intel, s már az 5. generációs chipjük, az EyeQ5 készül (13. *ábra*) melyet tovább fejlesztenek a vizuális információk szoftveres feldolgozásának hardveres támogatásában. A tervezett új generáció fogyasztása viszont várhatóan 10W-ra nő. [52]

A Renesas chip gyártó ARM alapú R-Car megoldása (14. *ábra*), melynek jelenlegi csúcsa a V3H model [53], a gyártó a megoldását az EyeQ5-t megelőző versenytársának tekinti. A Renesas megoldásában számos alapvető morfológiai művelet hardveresen támogatott, így a kétkamerás rendszerek támogatása, ehhez kapcsolódan diszparitás számolása, objektumosztályozás, valamint direkt hardveres támogatás a ConvNet megoldásához (14. *ábra*: „CNN-IP”), vagy az optikai áramlás számításához. [54]

Ezen két chipgyártó alapján a mély tanuló algoritmusok használatának trendje egyértelműen nyomonkövethető, viszont a korai látást segítő algoritmusokkal a veszélyhelyzetek felismerése tovább támogatható.



14. *ábra* A Renesas ARM alapú megoldásának rendszer diagrammja [53]

3 Objektum közeledésének felismerése biológiai alapú algoritmus segítségével

Egyes feladatok megoldását biológiai alapon megközelíteni nem újkeletű elgondolás. Az ember korábban is számos alkalommal lesett el praktikákat az élővilágból. Gondolhatunk egy olyan nagyon egyszerű dologra, mint amilyen a tépózár [55] is, melyhez intuíciót a bogáncs felépítése és tulajdonsága adott. A számításelméletben is foglalkoznak biológia motiválta algoritmusokkal a legkülönbözőbb területeken, legyen az jelfeldolgozás, kommunikációs protokoll, irányítástechnika, robotika, adatfeldolgozás, vagy akár az előző fejekben bemutatott gépi látás megvalósítása [56], melyeknél a neurális hálózatokra, elvekre épülő megoldások használata mára gyakorlatilag megkerülhetetlenné vált.

3.1 Közeledő objektum problematikája

A vizuális észlelés jelentősége régóta ismert, amin belül a mélységérzékelés problémaköre kurrens téma [57], ahogy egy videófolyamon a különböző mozgások felismerése, illetve szemantikai osztályozása. A közeledő mozgás felismerése elementáris feladatosztálynak tekinthető, melyre a biológiai rendszerek is megoldásokat adnak [58] [59] [60], attól függően, hogy az objektum közeledés jelenségének érzékelése mit jelent az adott élőlény számára: navigációs lehetőséget, vagy a túlélés kapcsán valamilyen közeledő veszély jelzését. Az egerek esetében a lecsapó ragadozómadár jelenthet közeledő veszélyt, míg a sáskák, madarak esetében a gyors és hatékony térbeli navigáció során van szükség a közeledő objektumok felismerésére.

Állatok [61] [62] [63] [64], de emberi csecsemők [65] vizsgálata során is bizonyítékot találtak arra, hogy a közeledő objektumok észleléséhez kapcsolódóan vannak előkészített idegrendszeri válaszok. A csecsemők vizsgálata ebben a tekintetben különösen érdekes eredményeket mutat: bár az 1-2 hónapos csecsemők látása homályos [66], emiatt a tárgyak alakja számukra nem megkülönböztethető, mégis van tanulmány [65] amely rámutat, hogy a világ megismerése nélkül is védekező módon reagálnak a közeledő tárgyra. Ez összhangban van a 2.1 fejezetben bemutatott vaklátás észlelésével, s feltételezhető, hogy az ehhez kapcsolódó neurális hálózat működése a tárgyak sziluettjének méretváltásával számol és alapvetően nem támaszkodik például a binokuláris látásból fakadó diszparitás észlelésére [67], vagy a tárgyak és méreteik ismeretére, amely a térbeli látáshoz elengedhetetlen.

A biológiában a közeledő objektum érzékelésének fontossága tetten érhető azokban a kísérletekben, ahol bebizonyították, hogy a legkülönbözőbb irányú mozgások közül a legjobban a közeledő mozgás keltette fel a kísérletben résztvevő emberi megfigyelők érdeklődését [68] [69] [70].

Az ütközésselkerülés miatt a jelenség gyors felismerése fontos a robotikában és az autópárhuzamos iparban egyaránt. A biztonság növelését célzó ADAS megoldásokkal jellemzően valamilyen szenzor (RADAR, LIDAR) adataira támaszkodnak, de a csak kamera alapú olcsóbb megoldások fejlődése egyértelművé vált az utóbbi évek éles, sok szereplős versenyében. Egy tanulmány megállapította, hogy egy rendszer, amelyik képes fél másodperccel előre jelezni az ütközést, az 60%-al csökkentheti a baleset megtörténtét, míg egy másodperc esetén már 90%-al [71]. A jelenlegi objektumközeledést detektáló megoldások közül van, amelyiket vállaltan az emberi látórendszer inspirált, mint amilyen a Mobileye [72]. A Mobileye első megoldásairól annyit lehet tudni, hogy négyzetes objektumok detekciójára épültek, melyeket képkockánként többféle osztályozási metodikával is megközelítettek [71].

A 2010-es évek elején az osztályozási problémák voltak a kutatások fókuszában: például az SVM és az AdaBoost algoritmust, olyan tulajdonságleíró módszerekkel kombinálták, mint a HOG, Haar, Főkomponens analízis, vagy a Gábor féle tulajdonságkinyerés [73, 74]. A neurális háló alapú algoritmusok kevésbé voltak elterjedtek még, a ConvNet épp ezen évtized közepén kezdett érdemben betörni a párhuzamos számításra képes architektúrák elterjedésével.

A közeledő objektum érzékelés biológiai megközelítésének egyik alaputatását a sáska látórendszere alapján készült ütközésselkerülő megoldása jelenti, mely a robotikában is szélesebb körben ténylegesen elterjedt [75] [76], viszont az előrejelzés minőségében korlátos képességekkel rendelkezett.

Egy közeledő objektum észlelése álló helyzetben alapvetően egyszerűen megoldható feladatnak tekinthető, mivel ekkor egy felismert objektum sziluettjének növekedésének megváltozásával, a tárgy vizuális elmozdulásával kell számolni. Nem állóhelyzet esetén viszont a tárgyon kívül minden más is elmozdul a képsíkon, így komplexebb feladattá válik a közeledő objektum észlelése.

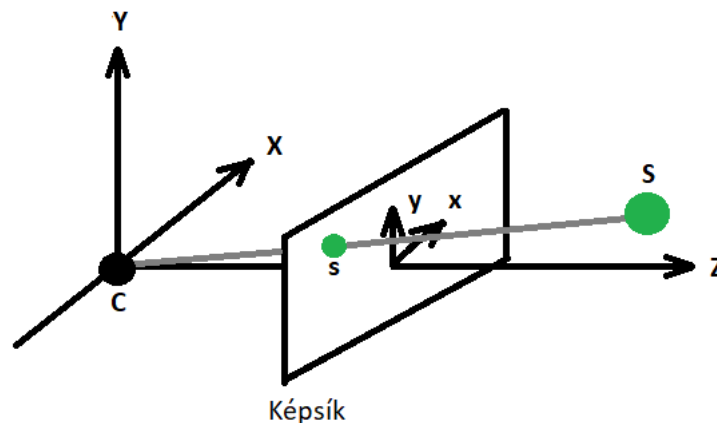
3.1.1 Térinformáció kinyerése

Normál esetben, a látás során, analóg, vagy digitális filmfelvétel készítésekor a tárgyakról visszaverődő fény valamilyen képsíkra vetül. A lyukkamera elvű képalkotást alapul véve

[77] egy $\mathbf{R}^3 \rightarrow \mathbf{P}^2$ leképezésről beszélhetünk (15. ábra), melyben \mathbf{R}^3 jelenti a három dimenziós projektív teret, melyben egy \mathbf{S} objektum leképeződik a \mathbf{P}^2 projektív síkra \mathbf{s} -ként, mely már nem tartalmaz térbeli információt.

A közeledő objektum felismerésének alapproblémája, hogy egy \mathbf{P}^2 projektív síkbeli képsorozat alapján hogyan lehet következtetni egy-egy objektum térbeli elhelyezkedésére. A középpontos projekció ugyanis csak annyit képes megmondani, hogy a \mathbf{P}^2 a tér melyik pontjainak leképeződése lehet, ami egy egyenessel írható le. Ilyen egyenes az $\mathbf{s}-\mathbf{S}$ -en keresztülhaladó vonal, ahogyan azt a 15. ábra is mutatja. Ez alapján belátható, hogy a képen szereplő objektumokról, vagy annak adott pontjáról sem méretbeli, sem távolságbeli információ alapvetően nincsen.

Felhasználva viszont, hogy a képalkotás során lencse rendszereket használnak fel, a geometriai optika alapegyenlete, a $\frac{1}{f} = \frac{1}{k} + \frac{1}{t}$ alapján a t tárgytávolság megadható, ha ismert az f fókusz távolság, valamint a k képtávolság. Erre az egyenletrendszerre építhető távolságbecslés az élességállítást felhasználva a digitális fotózásban. Másrészt a jövőben valamelyest segíthetnek a time-of-flight kamerák, melyek a mélység érzékelésre képesek. Ez a technológia manapság kezd kiforrni és terjedni a konzumer termékekben (pl. sorozatgyártott mobiltelefonokban való megjelenése napjainkban), s lehetőségei várhatóan túlmutathatnak majd az egyszerű lyukkamera modellen a térbeliség meghatározásánál.

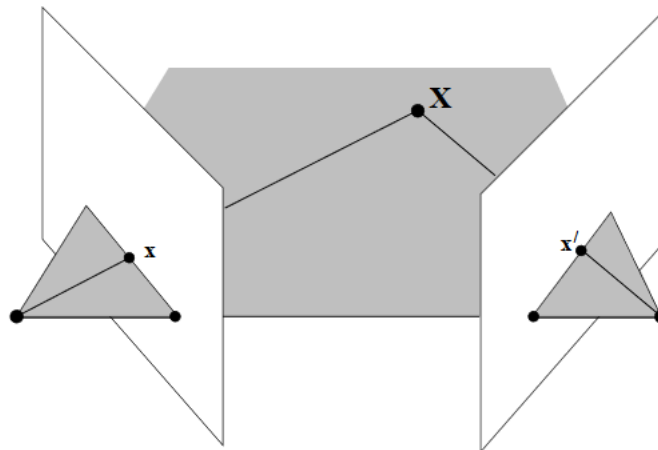


15. ábra A lyukkamera modellje: C a kamera középpont, S tárgy s pontra képződik le a képsíkon.

Az objektumok térbeliségének meghatározására több módszer is létezik (projektív rekonstrukciós eljárás több nézőpontból, 2, 3, vagy n nézőpont geometria, Euklideszi rekonstrukció, stb. [77]). Az előzőekben bemutatottak alapján az látható, hogy egy állóképen szereplő objektumok térbeli elhelyezkedéséről nem lehet pontos információt adni. Valamennyi előfeltételezéssel lehet élni, így például a kép egyes részeinek élessége alapján, vagy a kép értelmezésével, objektumok felismerésével és azok egymáshoz képesti arányával, ami már egy magasabb absztrakciós szint megjelenését is feltételezi.

A térbeli leképezés során az egymással párhuzamos vonalak a kép egy távoli pontjában futnak össze. A különböző párhuzamos egyenesek más-más pontokba tartanak össze, amit távlatpontnak nevezünk. A távlatpontok egy egyenesre fekszenek rá, amely egyenes megadja a kép horizontját [78] [28]. Ezzel az eljárással a kép perspektivikusságát felhasználva egyszerű feltételezést adhatunk a kép térbeliségére vonatkozóan, ami lehetőséget ad a tárgyak egymáshoz viszonyított helyzetének meghatározására a távlati vonalak mentén, amelyet már további előfeltételezésekkel lehet pontosítani (ismert méret, alak).

Legalább két különböző nézőpontból készült képpel már lehet előismeret nélkül is a térbeliségről információt adni, ekkor alkalmazhatóvá válnak olyan eljárások, amelyek epipoláris geometriai alapokon nyugszanak. Ilyen például a projektív rekonstrukció.



16. ábra Háromszögelés két nézőpontból, ahol x és x' pontok az egyes képsíkokon a leképezések. ([78] 10.1 ábrája)

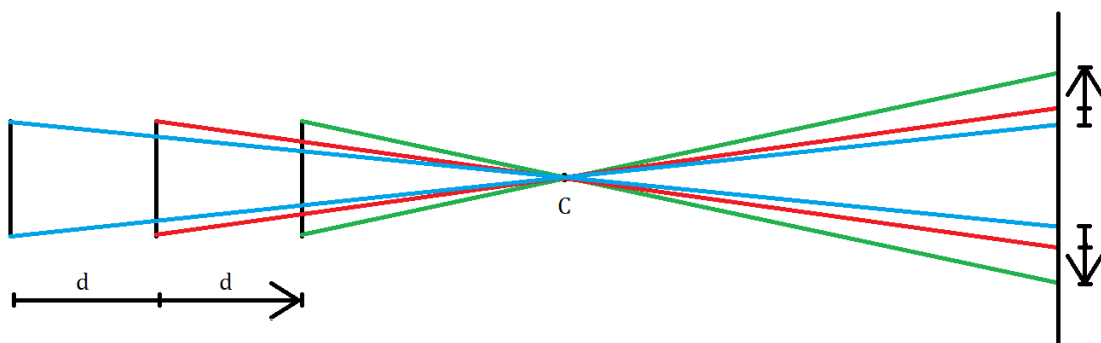
A 16. ábra bemutatja, hogy két nézőpont esetén a leképezés alapjául szolgáló \mathbf{X} pont hova esik a képsíkokon. Ezt az egymásnak való megfeleltetést az F fundamentális mátrix valósítja meg, ekkor teljesül az $x'Fx=0$ összefüggés. A fundamentális mátrix szinguláris érték felbontással, vagy a 8 pont algoritmussal számolható. A fundamentális mátrix ezt követően a projektív rekonstrukcióban felhasználható, melynél a kamerák számolhatók, mellyel már számolhatóak a térbeli pontok.

Álló platformon készült videófelvételnél felhasználhatók a képkockák közötti különbségek a térbeliség valamilyen szintű meghatározására. A képen való elmozdulások elemzése szolgálhat a térbeliségről információval (elmozdulás egymás elé-mögé, árnyékok megjelenése, eltűnése, változása). Forgó kamera mozgás során lehet a térbeli elhelyezkedésre következtetni, mivel a közelebbi tárgyak elmozdulása a képsíkon kisebb, viszont a legtöbb információhoz a kamera elmozdulása során jutunk. Ekkor az optikai áramlás elemzése már lehetőséget ad komolyabb következtetések levonására.

3.1.2 Közeledés a vizuális térben

A térbeliség érzékelésére a binokuláris diszparitás segítségével az élőlények néhány méter távolsáig képesek. Ez az embernél ez mindösszesen 30 métert jelenthet, de 9 méter fölött drasztikusan csökken a felismerés gyorsasága és pontossága. [67]

Az alábbiakban bemutatom, hogy a vizuális térben egy objektum közeledése a szemlélő számára milyen geometriai változást jelent. Egy közeledő objektum képének a képsíkra vetülő nagysága, vagyis a sziluettje a közeledés során folyamatosan nő. Két időpont közötti sziluettnek a mérete, a sebesség, a kép rögzítési sebessége és a képsíktól való távolság függvényében fog növekedni.



17. ábra A fókuszponthoz közeledve a képsíkra vetülő sziluett méret folyamatosan és nem egyenletes arányban növekszik.

Tekintsük egy homogén, egységnyi felületű tárgy képsíkra vetülését a C fókuszponton keresztül. Ha a tárgy elmozdul a képsík felé d távolságot, akkor annak sziluettváltozása az (1) szerint alakul egy C fókuszpontból tekintve.

$$A_{rel} = \frac{A}{d^2} \quad (1)$$

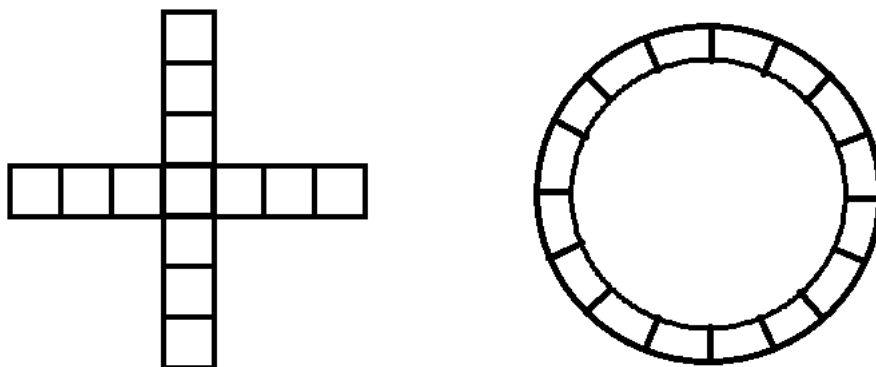
$$dA_{rel} = \frac{A_{rel}(t)}{A_{rel}(t-1)} = \frac{\frac{A}{d_t^2}}{\frac{A}{d_{t-1}^2}} = \frac{d_{t-1}^2}{d_t^2} = \frac{(d_t + \varepsilon)^2}{d_t^2} = \frac{d_t^2 + d_t \varepsilon + \varepsilon^2}{d_t^2} > \frac{d_t^2 + \varepsilon^2}{d_t^2} = 1 + \frac{\varepsilon^2}{d_t^2} > \frac{\varepsilon^2}{d_t^2} \quad (2)$$

A_{rel} határozza meg, az A objektumterület C fókuszponttól vett távolságtól függő területváltozás arányát, mely négyzetes összefüggést mutat. Ha egységnyi időben történt változást nézzük, akkor a közeledés területarányváltozásának mértéke megadható a (2) szerint két egymást követő időpillanat vett változásával.

A területváltozás nagysága így a fókuszponttól való távolság és a közeledés sebességének függvénye lesz, amely alapján a távolság-sebesség-területváltozás hármából kettőt ismerve a harmadik megadható. Az alapfeladat megoldható a növekedés tényének felismerése és a sziluett változás által.

3.1.3 Közeledés érzékelésének klasszikus algoritmikus módszerei

A legegyszerűbb közeledő mozgást érzékelő algoritmusok [79] alapvetően támaszkodnak az objektumszegmentálásra, illetve az élkeresésre. Ezek az egyszerű algoritmusok a területnövekedésnél a sziluett változás radiális irányú változását detektálják, amihez valamilyen detektor alakzatokat használnak, pl. keresztet vagy kört tipikusan (18. ábra).

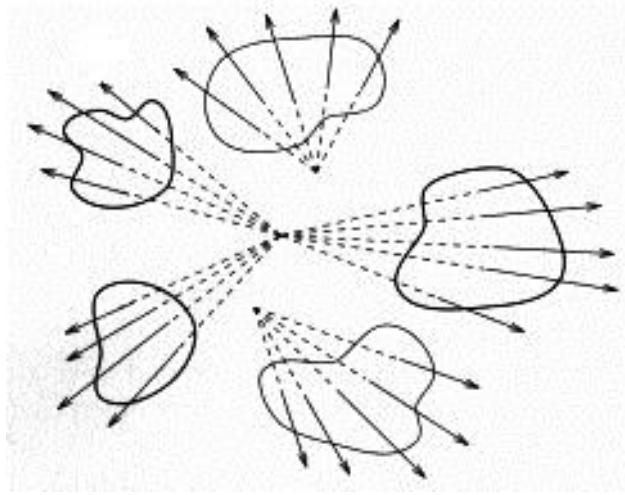


18. ábra Szimmetrikus detektorok sziluettnövekedés érzékelésére [79]

Ezek a detektorok a szabálytalan alakzatokkal nehezen birkóznak meg. A szabálytalan alakzatok felismerése a területalapú számításokkal és trajektória vizsgálattal lehetséges, mivel az egyes képkockák között a megtalált objektumokat meg kell egymással feleltetni és ezekre lehet területváltozást számoltatni.

További lehetséges detektálási eljárást ír le a [80] megoldás, mely egy referenciaközpontra támaszkodik és ez alapján próbál következtetni a közeledő mozgás irányára. A növekedés tényét jelző térinformáció, a közeledési fókusz hely meghatározása (FOE – Focus of Expansion), a nem kamerásíkbán történő közeledés esetén ad információt. Ez a szemlélő számára egy végtelen távoli pont irányából történő növekedést jelent az egyes képrészletek, objektumok esetén. A FOE alapú detektálás lényegét tekintve az optikai áramlás jelenségek felvételeken való felismerésén alapul.

Előre haladó mozgásnál ez alapvetően egyenletes a képtérben, laterális mozgás esetén, vagy a képhez képest gyorsabb, vagy lassabb növekedés esetén lehet következtetni a képrészletekben fellelhető eltérő működésre, mozgásjelenségekre. Az elvet a 19. ábra mutatja be.



19. ábra Focus of Expansion, vagyis a növekedés fókuszpontja ([80]-ból átvett).

Mozgás közben a kamera felé irányuló, a szokásosnál gyorsabb optikai áramlás utalhat közeledésre. Objektum azonos sebességgel való követésekor, pl. egy autónál, az egyforma sebesség miatt az autó képénél nem lesz megfigyelhető optikai áramlás, míg gyorsításakor vagy fékezésakor igen, mely a térben nagyobb változást okoz. A változás mértékének ismeretében pedig felírható következtetés az ütközés bekövetkezésének várható idejére. Napjainkban is számos kutatás épül [81], melyek már erősen építenek a ConvNetre, annak hatékony objektumfelismerési képességére is [82], ugyanis ezzel a megközelítéssel hatékonyan jelölhető ki azok a területek, amelyeknél az optikai áramlást vizsgálni kell.

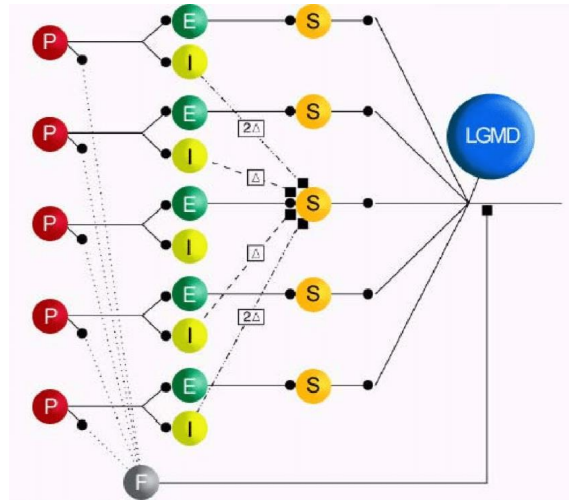
3.1.4 Közeledés érzékelésének biológia motiválta módszerei

Az objektumok közeledésének érzékelésének felismerése biológiai alapon nem újkeletű [83] [84] [75], a legszélesebb körű kutatások a sáskák (*Locust*), ecetmuslicák (*Drosophila*) [85] [86] [87] és a galambok esetében történtek [6]. Néhány élőlény esetében pedig már megtalálták azon sejtcsoportokat, mely a közeledés érzékeléséért felelős. Az egyes megoldásokból az látszik, hogy a természet is a geometriai növekedés felismerését veszi alapul, de az egyes élőlényekben ezt különböző módon teszi, mely az életmódtól is függ. Az alábbi fejezetekben áttekintésre kerül néhány megoldás, köztük az első emlősből felfedezett retinacsatorna is. A megoldások közös jellemzője, hogy gyorsan képesek a közeledés tényéről visszajelzést adni.

3.1.4.1 Sáskákban található sejtcsoport bemutatása

A legtöbb objektum közeledés érzékeléssel kapcsolatos kutatás a sáskák köré csoportosult, mivel a sáskákban 1974-ben fedezte fel O'Shea, Rowell és Williams a mozgás felismeréséért felelős látóneuront, a „*Lobula Giant Movement Detector*” (LGMD1) [59]. Ez a neuron az *ommatidium*mal áll kapcsolatban, mely a 2.2.1 *A látórendszer működése* fejezetben ismerttetett módon fogadja a külvilágból származó vizuális ingereket. Az LGMD1 mellett a feldolgozásban még jelentős a szerepe a „*Descending Contralateral Movement Detector*”-nak (DCMD) is, mely a kisebb objektumok mozgására reagál [88].

Az LGMD1-ből egy pár van a sáska látórendszerében, ahogy azt a 20. ábra bemutatja, amely folyamatosan dolgozza fel az egyes *ommatidium*ok irányából érkező képet, mint ingert (**P**). A több lépcsős, párhuzamos feldolgozás során az első lépésben a serkentő (**E**) és gátló (**I**) neuronok dolgozzák fel a vizuális ingert. Az egyes gátlóutak neuronjai, a szomszédos serkentő utakhoz kapcsolódóan fejtenek ki gátlást, ezt összegzi az LGMD1.



20. ábra Sáska retina modelljében az LGMD1 helye ([89])

A sáska esetében a modellben jelenlévő szomszédos gátlás feltűnő működési érdekesség, mely adódhat abból, hogy a sáska alapvetően kiakar térni menet közben a szembejövő madarak elől. Emiatt a kettősség miatt a felismerés tárgyát is gyorsabban kell detektálni menet közben, így a működése a navigáció támogatására is felkészített. Matematikailag a működés a (3)-as képlettel írható le [90].

$$v_i(t+1) = P_i \cdot v_i(t) + g_i^{exc} \sum_{j=1}^{N_{exc}} w_{ij} a_j(t - \delta_{ij}) - g_i^{inh} \sum_{k=1}^{N_{inh}} w_{ik} a_k(t - \delta_{ik}) \quad (3)$$

ahol P_i a membránpotenciált jellemző érték, g_i^{exc} és g_i^{inh} pedig az egyes serkentő és gátló csatornák erősítései, w_{ik} pedig meghatározza a szinaptikus, neuronok közötti kapcsolat erősségét i és valamely k neuron között, amit egy δ_{ik} késleltetéssel kell venni.

A neuron a_i aktivitását a t időpillanatban pedig a (4)-es képlet adja meg.

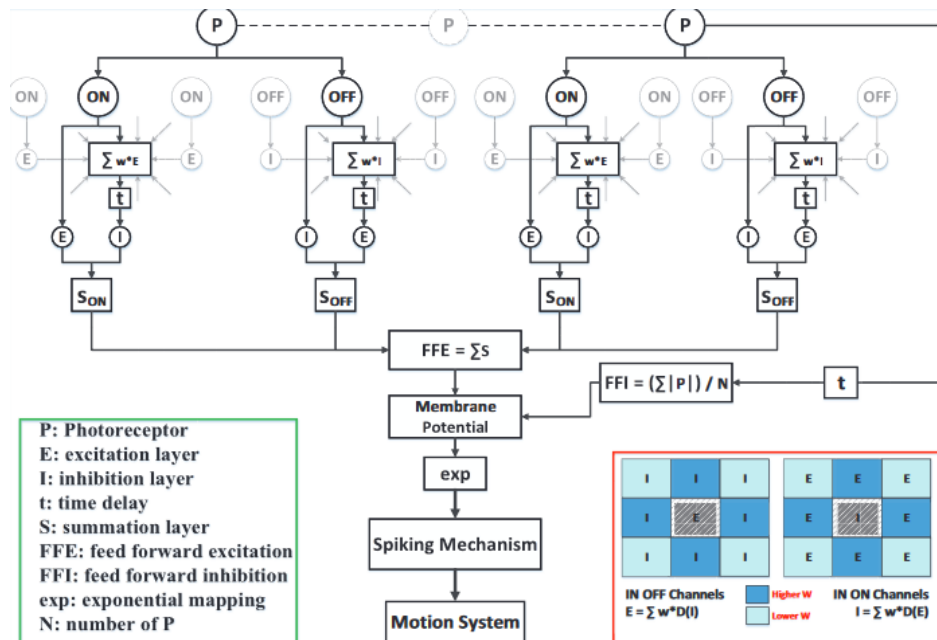
$$a_i(t) = \begin{cases} v_i(t), & \text{ha } v_i(t) \geq \theta \\ 0 & \end{cases} \quad (4)$$

Az 20. ábra bemutat egy előrecsatoló szakaszt is (**F**), mely a nagymértékű változások összegzésekor lesz aktív, például nagy területű mozgások esetén, amikor globális gátlásra van szükség. Ez az afferens lehetővé teszi, hogy a vizuális stimulus lefordítása nélkül reagáljon gyorsan a változásra, így a nagyarányú közeledő mozgás felismerésére. Az (5) képlet ezen utat számolja, melyben az α az aktuális időpillanatban az előre csatolás súlya, míg a β adja meg a pillanatnyi neuronaktivitás előre csatolás súlyát. Ha az F meghalad egy meghatározott küszöbértéket, akkor aktíválja a kimenetet.

$$F(t + 1) = \alpha \cdot F(t) + \beta \cdot \sum_{i=1}^N P_i(t) \quad (5)$$

Az LGMD1 modellel való kísérletek úgy találták, hogy a modell alkalmas az ütközés elkerülésére, több kutatási is kiegészítette az alapmodellt különböző optimalizálásokkal a még jobb eredmények elérése érdekében. A modell olyannyira sikeresé vált, hogy integrált áramköri implementációja is készült [83] [84] [75], valamint érzékelő processzortöm-bön való implementációt is kapott [90].

Az LGMD1 után az LGMD2 neuront Rind 1987-ben azonosította [91], viszont közeledésre való érzékeny működésének karakterizálása 1997-ben lett leírva [92]. Az LGMD2 a karakterisztikája alapján sokkal inkább volt alkalmas time-to-collision számolásra, mivel a tüzelési mintázata a tárgy közeledésével sokkal arányosabb volt az LGMD1-hez képest. A neuron számítási modelljét 2015-ben publikálták [93]. Az LGMD2 az LGMD1-hez hasonlóan épül fel, ugyanúgy tartalmazza az előre csatolt utat, viszont a gátló és serkentő utak számolásánál az OFF és ON csatornák számolásában már különbség van [94], mivel mindkettő gátlásával és serkentésével számol a modell, ezt a 21. ábra mutatja be.



21. ábra Az LGMD2 neuron modellje ([94] ábrája átvéve)

3.1.4.2 Galambokban található sejtcsoport bemutatása

A madarak esetében a galamboknál találták meg azokat a neuron afferenseket [64] [95], melyek az ütközés elkerülésben szerepet játszanak. Több kutatás is foglalkozott az ütközésig bekövetkező idő (time-to-collision) számításával, melyek a galamb agyának *nucleus rotundus* területéhez kapcsolódnak és a megjelenő aktivitásmintázattal foglalkoztak [60]

[58] [96] [97]. A modell közeledésre adott válasza a sáskáénál finomabb válaszra ad lehetőséget és legalább hasonlóan alkalmas választ képes adni, mint a sáska látórendszere [60]. A galamb látórendszer kapcsán agyi terület aktivitás vizsgálatáról írtak eddig tanulmányokat, a retinájában az nRT tektális neuronok felelnek a közeledés érzékeléséért [98], de ezekhez kapcsolódó számítógépes modellkísérlet nem kutatta a működés számítási hátterét, viszont viselkedése az LGMD2-höz hasonlít. A tektális neuronokból három félért azonosítottak (Tau, Rho, Eta), különböző érzékenységi görbékkel jellemezhető működésük.

3.1.4.3 Az egérben található sejtcsoport bemutatása

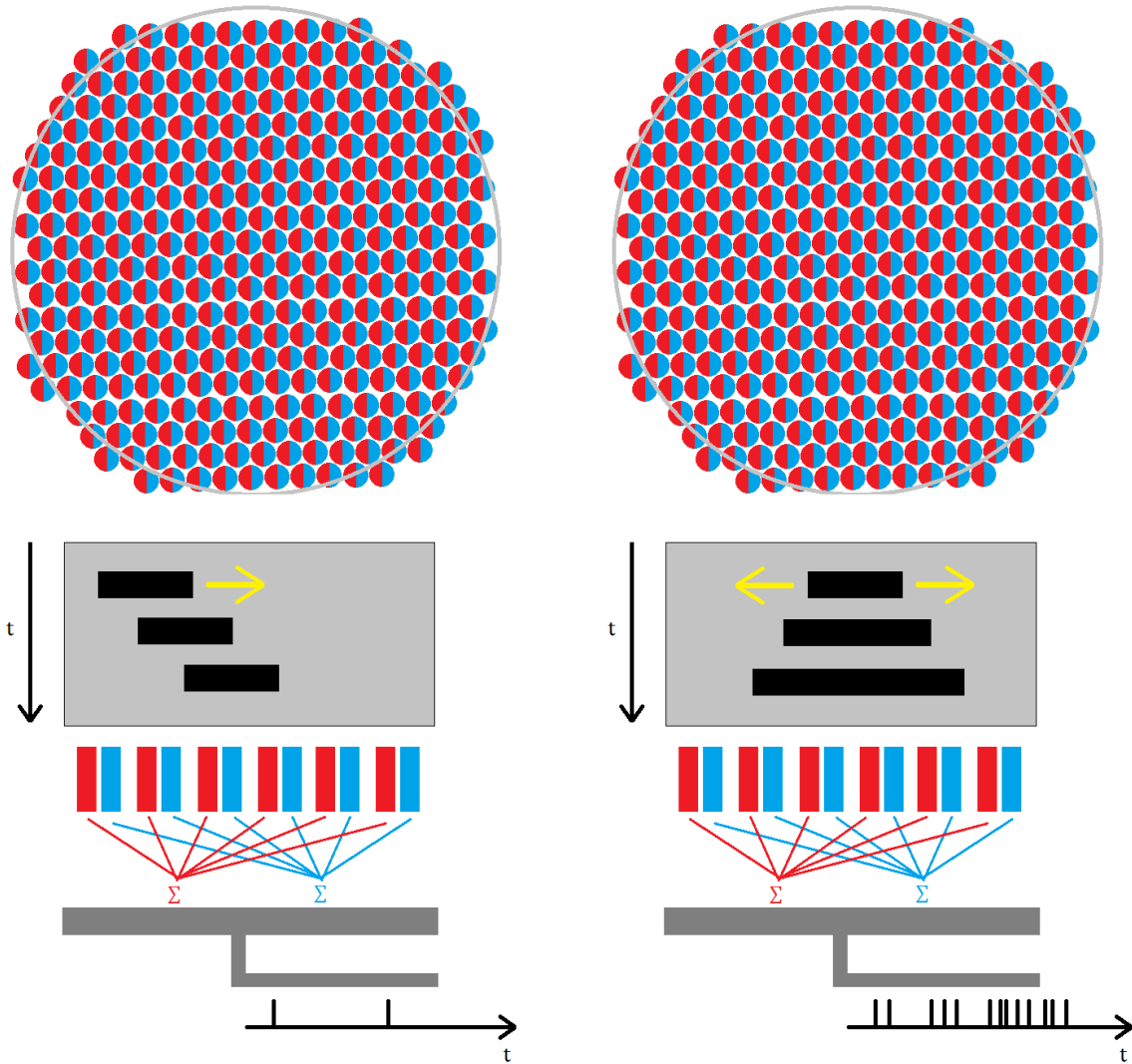
Az első emlős, melynek retinájában megtalálták a közeledés érzékeléséért felelős ganglionsejtet, az az egér volt. Az eredményt Roska Botond svájci kutatócsoportja publikálta 2009-ben [6].

Az egérben megtalálható, közeledésre érzékeny neuron, melyet sikerült genetikailag azonosítani, a Pvalb-5 ganglionsejt (vagy röviden PV-5), mely erősen aktivált állapotba az objektumközeledésre reagál, míg minden más jellegű mozgás elnyomja.

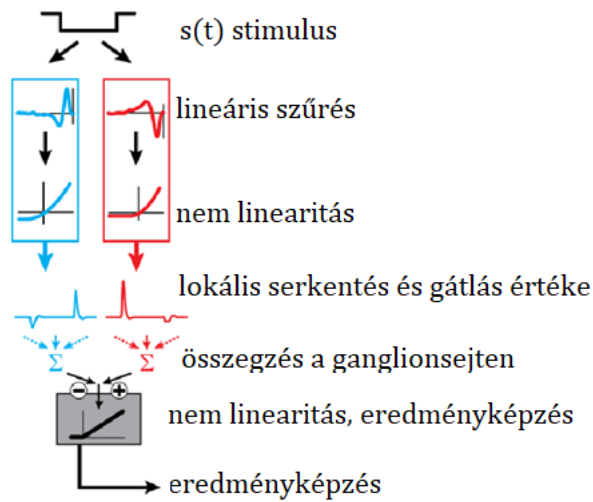
A sejt által adott válasz a már korábban felfedezett, hasonló célú látó neurális válaszokhoz képest kissé eltér. Ennek egyik oka lehet, hogy a sáskákhoz vagy a madarakhoz képest az emlősök mozgása lassabb, így másfajta védekező mechanizmusokra van szükség, így a működésük kapcsán a navigációs szempontok kevésbé érvényesülnek.

A Pvalb-5 ganglionsejt egy OFF típusú ganglionsejt, amelynek a negatív irányú változások (világosból sötétbe változás) serkentik az aktivitását. Kísérletileg igazolták azt is, hogy a sejt ON irányban erősen gátló. Ez a kettősség biztosítja azt, hogy a ganglionsejt a laterális mozgásokra nem reagál. Egy az állat felé mozgó sötét objektum sziluettjére az OFF típusú serkentő ganglionsejtek aktivitással reagálnak, míg az ugyanazon receptívmező területéről a laterális mozgás miatt a párhuzamosan fellépő erőteljes gátlás kialszítja a választ.

Egy Pvalb-5 ganglionsejt a látómező mintegy 10° fokát fogja át. A ganglionsejtek elhelyezkedése átlapolódó. A 10° -os látómezőt tekintve, egy 0,9 méter átmérőjű objektumnak, minimum 5 méterre kell elhelyezkednie, hogy még éppen látható legyen és elférjen a ganglionsejt által lefedett receptívmező területében, vagyis egy sejt ennél közelebbi érzékelést nem tesz lehetővé. Ennek élettani vonatkozása, hogy egy ragadozómadár, mint például egy parlagisas támadáskor, egy 60-80 cm átmérőjű, vagyis ez a látómező méret elég lehet támadás kellő időben történő észleléséhez és a támadás irányának lokalizálásához.



22. ábra Pvalb-5 ganglionsejt által lefedett receptívmező. Bal oldalon laterális mozgás esete, jobb oldalon közeledő mozgás (sziluett növekedés esete), melynél a serkentő és a gátló válaszok a ganglionsejten összegződnek.



23. ábra Pvalb-5 ganglionsejt számítási modellje valamely beérkező $s(t)$ stimulusra.

Egy Pvalb-5 ganglionsejt nagymennyiségű gátló és serkentő bipoláris sejttel áll kapcsolatban a receptívmezőn. A receptívmezőt érő stimulusokra adott sejtválaszokat vizsgálták, így a laterális mozgásokat, közeledő és távolodó mozgásokat. Ezekre a ganglionsejt különböző képpen adott tüzelő választ. A 22. ábra bal oldalán látható laterális mozgás esetén ritkán volt sejtaktivitás, míg a sziluettnövekedés esetén a növekedés mértékével arányosan növekvő, tüzelő válasz keletkeztek. Ezen modell szerinti működést mutatja be a 23. ábra: a beérkező stimulus alapján párhuzamosan egy serkentő és egy gátló csatorna segítségével, párhuzamosan számolható a sejtválasz. A két csatornán a lokális stimulus időbeli lefolyása konvolválva van egy lineáris szűrővel, ami a sejtválasz időbeli lecsengésével van összefüggésben. Ennek eredménye egy nemlineáris egyenirányításon megy át, mely a lokális gátlások és lokális serkentések erősségét mutatja, s melyek a receptívmezőt alkotó ganglionsejtre stimulusként hatnak. A ganglionsejt ezen sejtválaszokat összegzi, mely végén ismét egy nemlinearitáson keresztül adódik a ganglionsejt válasza, mely a tüzelés nagyságának feleltethető meg.

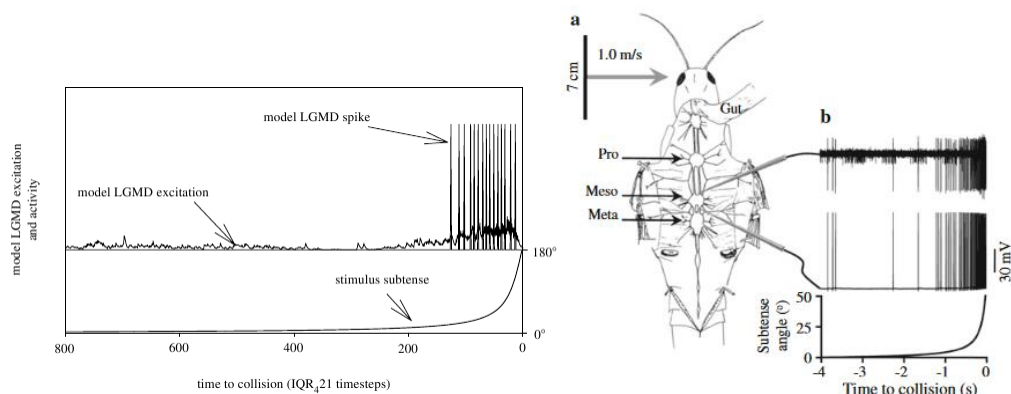
3.1.4.4 Biológiai modellek összehasonlítása

Az emlős és a rovar közeledő mozgását feldolgozó sejtcsoportok több ponton is különböznek, viszont alapvetően a működési elv, vagyis a sziluettt változás vizsgálata hasonló [85]. A fő különbséget a modellek között a rovarok és az emlősök között lévő szemfelépítés jelenti, mely más-más megoldási variációra ad lehetőséget, így jelenhet meg a rovarok esetében az előreccsatolás a válaszok korlátozására, mely az emlősök esetében az átfedő receptívmezőkkel képes megvalósulni.

A rovarmodellben a gátlás kisebb területen jelenik meg és közvetlenül a serkentő csatornára próbál hatni, míg az emlősnél a gátlások a receptívmező bármely területéről a modell egészére hatással vannak és csak az OFF irányban hatnak.

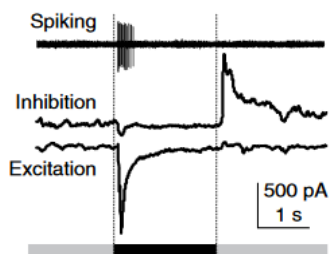
A rovar látórendszerének kompaktsága miatt a feldolgozás sebességében lehetnek különbségek. Fontos különbség, hogy a sáskánál az előreccsatolt gátló mechanizmus miatt egy nagyobb változás automatikusan sejtválaszt vált ki, valamint ez a válasz nem nő együtt a stimulus növekedésével, amelyet a közeledő objektum területváltozása okoz. Az egérnél a válasz a stimulussal arányosan változik. Az 24. ábra alul látható a stimulus nagysága, ami a területváltozással egyezik meg. A modell LGMD1-n mérhető válasza („model LGMD excitation”) láthatóan nem követi a stimulus nagyságát. Végeztek mérés-

seket az LGMD1 posztszinaptikus párján a DCMD-n is. Ezen a retina felé egy 7cm átmérőjű fekete korongot 1m/s-al közelítve, kb. 1 másodperccel korábban kezdett el tüzelni.



24. ábra Bal oldalon az LGMD1 válasz a közeledésre - alul a retinára eső sziluett nagysága adott időpillanatban, felül pedig az LGMD1 sejt válasza ([58]-ból átvéve). Jobb oldalon a DCMD válasza 1m/s közeledésre ([83]-ból átvéve).

[99] Az LGMD2 sejtválasza viszont a biológiai kísérletek alapján a közeledő objektum sziluettváltozásához hasonló erősséggel tüzelt.



25. ábra 400 μm -es fekete objektum 2 mp-ig történő megjelenésére (vonalak között) adott Pvalb-5 sejtválasz. ([6]-ból átvéve)

A Pvalb-5 ganglionsejt válasza a stimulussal megegyező ütemben kezd nőni a közeledésre, vagy annak lassulására csökkenni (25. ábra). Ez lényegében azt is jelenti, hogy az emlős látórendszere kifinomultabb, jobban paraméterezhető, mert nem csak bináris, „igen-nem” jellegű szignálok adására képes, illetve erősebb jelzésre. Ez adódhat a retina finomabb felépítéséből.

Tekintettel arra, hogy a rovar retinában az LGMD alapvetően egy hálózat, addig a Pvalb-5 több átfedő, nem az egész látóteret egyként kezelő receptívmezőt jelent. Ennek megfelelően a time-to-collision a rovar esetében könnyebben számolható.

3.2 Biológia motivált közeledő mozgás érzékelés számítógépes modellje

A *3.1 Közeledő objektum problematikája* fejezetben látható volt, hogy a térbeliség hogyan számolható algoritmikus úton, valamint áttekintésre került a sáskák, galambok és a kutatásom idején frissen felfedezett, egerek látórendszerének működése a közeledő objektum érzékelése során.

Ebben a fejezetben áttekintem a széleskörben már elterjedt sáska látórendszer alapján készült megoldásokat, valamint bemutatom az általam, az egér retinája alapján készített modellt.

3.2.1 Ütközés elkerülési megoldás a sáska látórendszere alapján

A sáska és a galamb látórendszerében megtalált sejtcsoportok modellje alapján már lehetőség volt különböző ütközéselkerülő algoritmusokat fejleszteni. [100] A modellre épült egyik kutatás [75] a Volvo XC90 kereskedelmi forgalomban is kapható gépkocsi kapcsán történt ipari kutatás. A chip implementációját teljes egészében a sáska látórendszer által motiválva tervezték meg ezen vizuális képfeldolgozási feladatra. A rendszer implementálása kapcsán a külön szenzor és külön processzor helyett egy integrált megoldás megvalósítása volt a törekvés, annak számos előnye miatt (kiolvasás utáni adatátvitel egyszerűsítése, HDR képesség kihasználása). Már ezen megoldás során felmerült a pixelenkénti adatfeldolgozás, vagyis a feladat megoldása fókuszsíkbeli processzortömbbel.

Az ismertetett ütközéselkerülési megoldás alapvetően a sáska modell számításaira épít, s a közölt eredmények szerint képes volt 1 másodperccel az ütközés előtt a rendszer jelezni. A megoldás egyik alapvető konklúziója volt, hogy érdemes a klasszikus top-down fejlesztési metódusok mellé a biológiai megoldásokat kombinálni.

3.2.2 Egér retina alapú modellezés

Mint láttuk, a 3.1.4.3 fejezetben bemutatott Pvalb-5 ganglionsejt csoport működése jelentősen eltér a sáska látórendszerében található sejtcsoportok működésétől, valamint az egér életmódjából adódóan a látórendszere más célfeladatra fejlődött ki, sokkal inkább a támadó jellegű észleléseket részesíti előnyben a navigáció helyett. Az egér látópályája

mindemellett sokkal inkább hasonlít az emberére is. Emiatt is érdekes felvetésnek ígérkezett a biológiai modell működésének számítógépes modell segítségével történő validációja.

3.2.2.1 A Pvalb-5 sejtcsontra kvantitatív modellezése

A retinában folyó sejt szintű tér-időbeli jelfeldolgozás levezethető matematikai egyenletekkel melyet az időben folytonosnak, míg a térben diszkrétnek kell tekinteni. A levezetés egyes lépéseinek megértéséhez a 23. ábra nyújt segítséget.

A modell implementálásához használt Eye-RIS rendszer ugyan sok műveletet képes analóg módon megoldani, viszont az egyes lépéseket csak diszkrét módon lehet vele kezelni, így a modellezésnél mindenképpen szükség volt időben is diszkrét lépéseket megadni. Az alábbi egyenletek megadják a működés diszkrét matematikai modelljét, melyek a korábban bemutatott retinacsatorna mérési eredményeivel összevethetőek.

A modell bemenetén az optikai érzékelőről érkező fényintenzitás, mint $s(t)$ stimulus jelenik meg, míg a modell kimenete a Pvalb-5 ganlion sejt tüzelés nagyságának szintjét adja meg.

A matematikai modell első lépésként egy időbeli szűrést végez egy pixel bemenetén: a szűrés két párhuzamos csatornát állít elő, egy serkentő e - (6) és egy gátló i - (7) csatornát:

$$e_{i,j}(t) = \sum_{n=0}^{g-1} s_{i,j}(t-n)w_n^e \quad (6)$$

$$i_{i,j}(t) = \sum_{n=0}^{g-1} s_{i,j}(t-n)w_n^i \quad (7)$$

ahol:

$s_{i,j}$ a fotóreceptort ért fényintenzitás értéke, a képkocka i,j pontján, a modell bemenete,

g a diszkrét képfelvételek száma, mely az időbeli konvolúcióban résztvevő képkockák száma,

w_n^e a serkentő csatorna időbeli konvolúciójához tartozó súlytényezős vektor,

w_n^i a gátló csatorna időbeli konvolúciójához tartozó súlytényezős vektor,

$e_{i,j}(t)$ adja meg a serkentő csatorna t időpillanatban vett időbeli konvolúciójának eredményét az i,j képkockára (a szűrés serkentő kimenete),

$i_{i,j}(t)$ adja meg a gátló csatorna t időpillanatban vett időbeli konvolúciójának eredményét az i,j képkockára (a szűrés gátló kimenete).

Az g -el jelölt képfelvételek száma a modellben a modell azon tulajdonságát jelenti, hogy a retinára eső s fényváltozásnak tranziense, memóriája van a feldolgozás során.

Az időbeli konvolúciót egy nemlineáris átviteli függvény követi, melynek neurobiológiai szempontból van jelentősége. Egyrészt ez helyettesíti a biológiai modell azon tulajdonságát, miszerint a sejtek közötti kommunikációs csatornák unipolárisak. Másrészt a feldolgozási szempontból képes a kis változásokat elnyomni mind a serkentő, mind a gátló csatorna oldaláról. A nulla körüli értékek időbeli zajból származhatnak, vagy nagyon lassú intenzitás változás eredményeként jöhetnek létre, melyeket el kell nyomni a térbeli átlagolás előtt, hogy csökkenthetőek legyenek a hibás pozitív riasztási eredmények. A matematikai modell tekinthető egy egyszerű, egy töréspontos szakaszosan lineáris függvénynek, mely így egy nemlineáris karakterisztikát ad meg, ez pedig képes tükrözni a biológiai modellben is megtalálható működést. Az egy töréspontos függvényt a modellben az alábbiak lehet feltételezni:

$$h_e(x) = \begin{cases} (x+o_e), & \text{ha } (x+o_e) > 0 \\ 0 & \text{egyébként} \end{cases} \quad (8)$$

$$h_i(x) = \begin{cases} (x+o_i), & \text{ha } (x+o_i) > 0 \\ 0 & \text{egyébként} \end{cases} \quad (9)$$

ahol:

x : $e_{i,j}(t)$, vagy $i_{i,j}(t)$,

o_e a serkentő csatorna eltolási értéke,

o_i a gátló csatorna eltolási értéke,

h_e a gátló csatorna átviteli függvénye,

h_i a serkentő csatorna átviteli függvénye.

A csatorna kimenete a fenti függvénybe helyettesítve $h_e(e_{i,j}(t))$ és $h_i(i_{i,j}(t))$ -ként alakul, melyek a Pvalb-5 ganglionsejt bemeneteként jelennek már meg, amely a sejten összeadódik. A sejt kimenete pedig az alábbiaként alakul:

$$g_{k,l} = r \left(\sum_{s,t \in N_r(k,l)} (h_e(e_{s,t}) - h_i(e_{s,t})) \right) \quad (10)$$

ahol:

$N_r(k,l)$ a ganglionsejthez tartozó (k,l) méretű receptívmező,

$r(x)$ a sejtválaszt adó függvény, mely $r(x) = \frac{x(\text{sign}(x)+1)}{2}$ -ként írható fel.

A következőkben bemutatom a matematikai modell implementációját és elemzését, valamint karakterizálását.

3.2.2.2 A modell implementálása

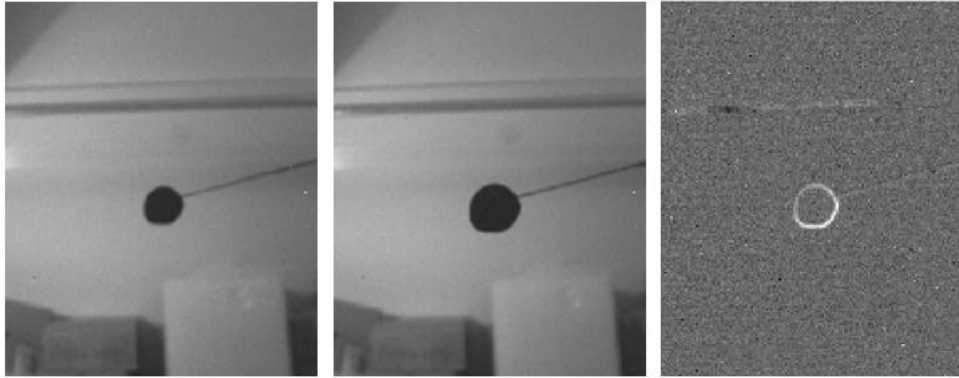
A modell implementálását analóg és digitális architektúrán egyaránt fontosnak tartottam kidolgozni. Az analóg implementációhoz az Anafocus cég Eye-RIS v1.2 [5], majd a stabilabb v1.3 rendszerét használtam, melytől azt vártam, hogy a modell a valósághoz a lehető legközelebbi sebességgel képes működni, tekintettel arra is, hogy egy olyan érzékelő processzortömböt használtam, mellyel a képkészítés pillanatában el lehet már a pixelszintű műveleteket végezni.

A modell MATLAB-os környezetben történt implementációja részben a modell működéséről számszerűsíthetőbb adatokat tud adni az analóg implementációhoz képest, valamint lehetőséget ad a modell finomabb paraméterezésű, megismételhető vizsgálatára. A digitális rendszeren történő implementációnál az órajel és a processzorarchitektúra nagyban meghatározza az elérhető feldolgozási sebességet.

A kvantitatív modell első lépése a serkentő és gátló csatornák számítása – ahogy az a modell ismertetésénél látható volt –, amely egy időbeli konvolúcióval kezdődik (23. ábra). A modell konvolúciós kernelét különböző hosszú és súlyozású kernelekkel vizsgáltam. A biológiai modellen végzett mérési eredmények azt mutatták, hogy elég közel áll egymáshoz a serkentés és a gátlás karakterisztikája laterális mozgás esetén, így a modellvizsgálatban egymás inverzeinek tekinthető a két súlyozás karakterisztikája.

A modellt különböző méretű és súlyozású kernelekkel vizsgáltam, ennek alapján megállapítható volt, hogy már a legegyszerűbb, $[-\frac{1}{2}, -\frac{1}{2}, 1]$ súlyú kernel is megfelelően jó eredményhez vezetett, vagyis három egymást követő képkocka feldolgozásával már értékelhető eredményt kaptam. Ez összevethető a mérések 50 ms körülre adódó élfelfutásához tartozó, arányos képfelvétel készítési sebességével is. A kernel választás szempontjából

fontos kitétel, hogy zéró összegűnek kell lennie, mivel ez adja meg a kép intenzitás változását. Ez a lépés az Eye-RIS rendszeren 16 ms alatt végezhető el és jól megadja a sziluett változását (26. ábra).



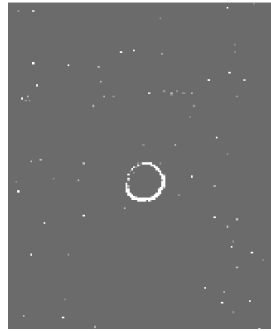
26. ábra Közeledő mozgáson pixelenként elvégzett időbeli konvolúció eredménye.

Nagyobb időbeli konvolúciós kernelt választva természetesen sokkal pontosabb eredmény kapható. A kernel nagyobbra vétele ugyanakkor a számítási komplexitás térbeli és adatátviteli követelményeire van hatással, valamint egyértelműen hatással van a rendszer dinamikájára is, tekintettel arra, hogy nagyobb időintervallum alapján számol változással. **A konvolúciós paraméter adja a modell első karakterisztikus paraméterét**, amely az egész rendszer hatékonyságára és reagálóképességére képes hatást gyakorolni.

Az analóg és digitális modell számítási komplexitásának szempontjából ez az első pont ahol komoly architektúrális előnyt jelent az analóg felépítés, mivel a három képkocka konvolúciója mindössze 16 μ s alatt elvégezhető a memória írásokkal és olvasásokkal együtt, ami az Eye-RIS 25 ezer képpontját (pixelét) egyenként véve, digitális architektúrán nem párhuzamosítva 1,5GHz-es órajellel történő műveletfeldolgozási sebességet jelent 176x144 nagyságú 8 bites képet tekintve.

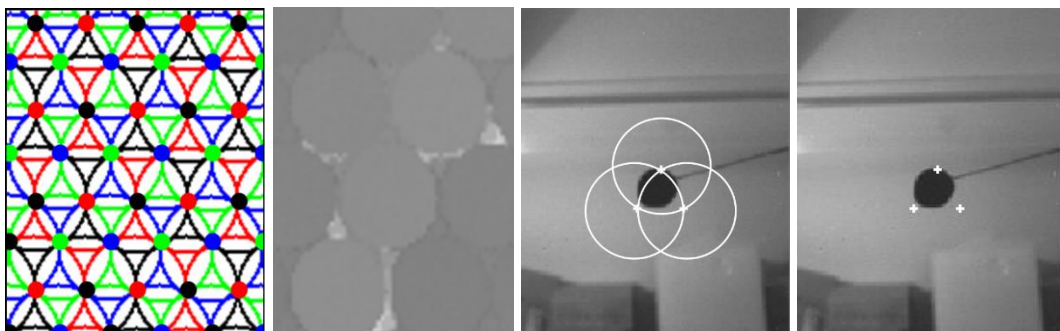
A felvázolt modellben a következő lépés a nemlinearitás megvalósítása, ami egy egy töréspontos, szakaszonként lineáris függvény számolását jelenti. A töréspont számolása egy eltolással vett 0-hoz képest vett küszöbérték számolás. **Az eltolási értékek meghatározása (o_e és o_i) adja a második karakterisztikus paraméterét** a rendszernek. Analóg műveletként ez a lépés 4 μ s alatt végezhető el az Eye-RIS rendszeren, mely digitális architektúrán már a 25 ezer képkockára vetítve 6,25 Ghz-es órajelet feltételez ezen a képméreten. A megfelelő offset beállításával alapvetően zajszűrés is történik, melyet a 27. ábra demonstrál. Az

eredmény képen az látható, hogy a 26. ábra jobb képének zajai elsimultak, elvétve található néhány zajos hely, vagyis sikerült elsimítani az analóg rendszerből származó zajt.



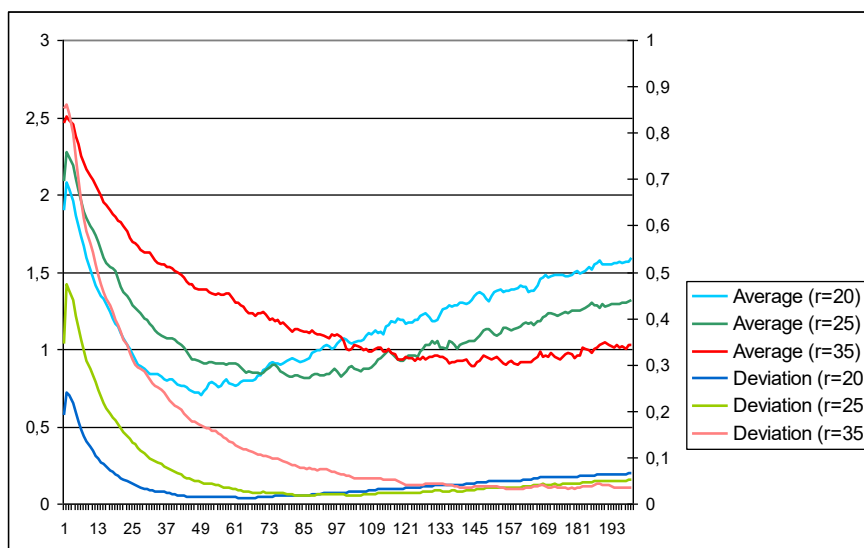
27. ábra Nemlinearitással számolt eredmény a csatornákon.

A modell harmadik lépésként a két csatorna térbeli összegzését végzi el. A művelet az analóg Eye-RIS rendszeren többféleképpen is implementálható. Alkalmazható eljárás a „ $mean()$ ” függvény az egész képre, vagy maszkolt képterületekre, mely így egy-egy receptívmezőt reprezentáló középértékét adja vissza. Emellett elvégezhető a képen a diffúzió művelet is, melyet követően a pixelértéket kiolvastva kapható meg a ganglionsejt összegzésének eredménye, ez sokkal közelebb áll a jelek terjedéséhez. Az analóg rendszeren a $mean()$ használata az Eye-RIS rendszeren 12 μs végezhető el, a diffúzió művelet pedig kalibrálható hosszúságú ideig tart. Ezen a rendszeren a kisebbre vett, maszkkal meghatározott receptívmezők esetén a diffúzió akár gyorsabb is lehet, míg az egész képet tekintve a $mean()$ függvény, az eredmények tekintetében mérhető különbség nem volt, figyelembe véve a rendszerben meglévő analóg zajt is. Átfedő receptívmezőket alkalmazva egy receptívmezőre a 28. ábra szerint látható módon számolódik a serkentő és gátló csatornák hatása.



28. ábra A bal oldalon látható átfedő receptívmezőket feltételezve, ganglionsejtenként a diffúziót elvégezve kapható eredmény. A b2 képen a terület szürkességi értéke a ganglionsejt aktivitásának meghatározója, a közeledés során pedig a b3 szerinti receptívmezők szerint várható válasz, ahogy a b4 képen is látható a kereszttek méretével arányosan.

Az átfedő receptívmezővel számolva láthatóvá válik, hogy a kép melyik részletén van a közeledés és mennyire intenzív, hány ganglionsejt jelez a térrészen sziluettnövekedést. A négyes, körös átfedést használva $50\mu\text{s}$ körül adódik a művelet elvégzése, amit szintén a 28. ábra mutat be. Az optimális diffúziós idők kiszámításánál a receptívmező méretében meghatározó (29. ábra), $r = 25$ px esetén az 50ms már 10% alatti eltérést jelentett.

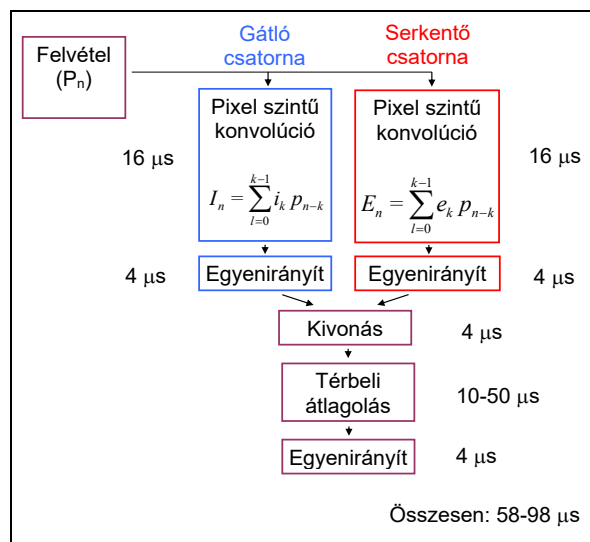


29. ábra Különböző diffúziós idők mellett (vízszintes tengelyen, ms) a diffúziós átlagolás képértékének (Average, bal tengely) alakulása különböző r sugarú receptívmező méretek esetén, eltérés a normál átlagoláshoz képest (Deviation, jobb tengely)

Többféle átmérőjű receptívmezővel is vizsgáltam a modellt, a gaussos diffúziós művelet $10\text{-}15\mu\text{s}$ alatt számolható az Eye-RIS architektúrán, ennek eredményét összehasonlítottam a $mean()$ függvényvel. Mivel a kalibrált környezetben a két eljárás közötti pontosságban nem lehetett különbséget tenni, így a modell szempontjából gyorsabban számolható és ahhoz közelebb is álló diffúzió lett a modell alapja. **Megállapítottam, hogy a modell harmadik karakterisztikus paramétere a receptívmező mérete**, mely hat a felismerhető objektumok nagyságára a modell érzékenységén keresztül.

A modell utolsó lépéseként történik a behelyettesítés, amely az előbbieket szerint számolt középérték eredményét vizsgálja: ha elér egy meghatározott küszöbértéket, akkor tekinti a rendszer úgy, hogy tüzel a ganglionsejt. **A küszöbérték alkalmas a modell érzékenységének kalibrációjára, ez a negyedik karakterisztikus paramétere a rendszernek.**

A 30. ábra folyamatábrája mutatja be modell számítását, melynek lépésenkénti eredményeit a 26. ábra, 27. ábra és a 28. ábra konkrét mérési eredményekkel mutat be. A modellben a serkentő és a gátló csatornákat párhuzamosan kell számolni. Tekintettel arra, hogy a hardverfelépítés általánosságban nem teszi lehetővé ezt az egyidejű párhuzamos számítást az Eye-RIS rendszeren sem, így csak azzal a közelítéssel lehet élni, hogy gyors egymásutánban, ugyanazon a képeken történik meg a számítás. Másik lehetőség, az az egyszerűsítés, hogy a két út konvolúciós ablakának súlyozását egymás -1-szereseinek tekintjük a számolásnál.



30. ábra Az implementált modell folyamatábrája, az analóg rendszer számára szükséges műveletek idejeivel

Az Eye-RIS modellben egy ganglionsejt tüzelésének számolása 58-98 μs közötti időben végezhető el, függően a korábban már említett receptívmezőt összegző függvényről, valamint a konvolúciós ablak méretétől.

3.2.2.3 A modell kalibrációja

Az implementálást követően első körben empirikus módon vizsgáltam meg a modellt, amely a cikkben leírtak szerinti válaszokat adta és alkalmas volt arra, hogy a modell implementációja ellenőrizhető legyen. A modell vizsgálatának következő lépéseként a modell kalibrációjával foglalkoztam. Ennek során megvizsgáltam a modell viselkedését meghatározó karakterisztikus paramétereket és határaikat.

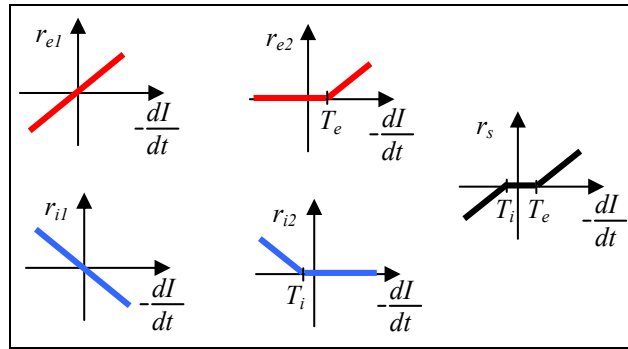
Időbeli konvolúciós ablak méret

Az időbeli konvolúciós ablak mérete (időbeli hossz), mint a rendszer első karakterisztikus paramétere hatással volt a rendszer válasz érzékenységre és tehetetlenségére, ahogyan azt előzetesen is vártam. A hosszabb ablak a modell érzékenységét növelte, valamint a feldolgozási időre is hatással volt, figyelembe véve egy-egy kép készítésének idejét és a képkockák alapján feldolgozandó változásokat.

Serkentő és gátló csatornák küszöbértékei

A serkentő és gátló csatornák küszöbértékének állításával a kisebb zajokból származó változások váltak kiküszöbölhetővé. A zaj részben a képkészítés során jelenik meg, másrészt az analóg feldolgozás során, mivel az analóg memória cellák az időben változnak az Eye-RIS rendszerben (az információt hordozó töltésmennyiségből folyamatosan veszítenek), valamint a külvilágból érkező zajokra is érzékenyek a rendszerben lévő elemek. Ez a biológiai rendszerekben is létező jelenség, így a modell kiértékelését nem befolyásolja. A számítások során ezek a zajok jól láthatóak (26. ábra képfelvételei láthatóan nem homogének), a küszöbérték pedig ezen zajhoz állítható be alapértelmezésként. A küszöbértékekkel kapcsolatos érzékenység összefüggéseit szemlélteti a 31. ábra is, melyen a vízszintes tengelyen a sötétség/világosság intenzitásváltozás látható időben, a konvolúciós ablakon belül feldolgozva. A serkentő (T_i) és a gátló (T_e) küszöbértékekkel a modell különböző karakterisztikájú zajokra konfigurálható. A jobb oldali feketével jelzett átmenet mutatja, hogy milyen kis intenzitás-változásokra nem érzékeny a rendszer.

A modell ezen a pontján megállapítható az is, hogy a laterális mozgásokkal kapcsolatos változásokra egy receptív mezőn belül ez nem lesz hatással, vagyis amikor az elmozdulás miatt ugyanakkora terület válik sötétté, mint világossá egy receptívmezőn belül. Itt hatással a háttér pl. gradiens jellegű különbségéből fakadó változás lehet, mellyel a küszöbérték beállításánál szintén számolni kell. Továbbá számolni kell azzal, hogy a tárgykép receptívmezőbe érkező, vagy onnan kilépő laterális mozgásai, vagy mozgás közben az optical flow jelenségek is ezen a ponton konfigurálandó zajként jelennek meg. Egy közeledő objektumnál egyértelműen a serkentő irányú intenzitás változások lesznek többségben.



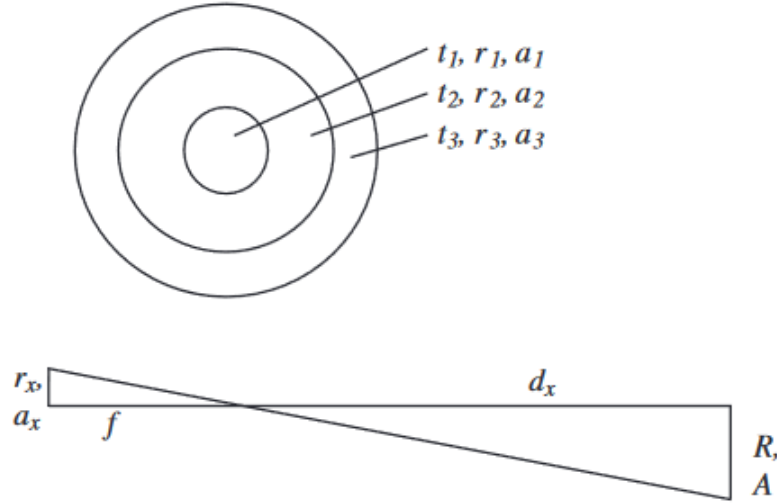
31. ábra A serkentő (piros) és gátló (kék) csatornák válasza az időben történő intenzitásváltozásra. A jobb oldali fekete mutatja az együttes választ, melynél a $T_i = x + o_i$ és $T_e = x + o_e$ értékek a csatorna küszöbértékeitől függenek.

Receptívmező mérete

A receptívmező mérete a látómező egy szeletét fedi le, így ezen paraméter megválasztása meghatározza, hogy mekkora objektumok közeledésének észlelésével számol a rendszer. Szűk receptívmező esetén vagy nagyobb méretű távoli objektum közeledése, vagy kisebb objektum korai észlelése lehetséges. Az apertúra probléma ezen a ponton jelenik meg a modellben, így az adott receptívmezőn túlnyúló sziluett a közeledéshez képest kisebb tüzelési értékkel jelenik meg, vagy a receptívmezőt meghaladó méretű sziluett esetén nem feltétlenül jön tüzelésbe a ganglionsejt. Ekkor a tárgy sziluettje más receptívmezőkön is észlelhető már, az átfedés sűrűsége a teljes érzékelő rendszer egészét tekintve további paraméterként szerepel. Magát viszont a megismert sejtmodell értékelését önmagában nem befolyásolja.

Ganglionsejt érzékenységének küszöbe

A rendszer végső érzékenységét befolyásoló negyedik paraméter a a ganglionsejtnak meghatározott küszöbérték. Ez a fentiek függvényében határozza meg, hogy a közeledés jelezhető-e, tüzelésben van-e a sejt. A válaszok alapján megállapítható, hogy a sötét közeledő objektumokra a rendszer egyértelmű választ képes adni, csak a ganglionsejt kimenetéből olvasott átlagértéket tekintve a sejtválasz a közeledő objektum mozgásával folyamatosan nő, míg laterális mozgás és távolodás esetén konstans, zajszerű.



32. ábra Közeledés sziluettváltozásához kapcsolódó értékek

A sejt válasznövekedésénél és a küszöbök meghatározásánál a konvolúciós ablakkal számolni kell. A korábbiakban említett három hosszú memóriát feltételezve a válasznövekedés az alábbiak szerint számolható:

Tekintsük a $[1, -1/2, -1/2]$ ablakot. Ekkor a konvolúciós ablak súlyait behelyettesítve az alábbiak szerint alakul a ganglion sejt válasza:

$$g_{k,l}(t) = h_g \left(\sum_{i,j \in N_r(k,l)} \left(h_e \left(\sum_{n=0}^s I_{i,j}(t - n\Delta t) w_n^e \right) - h_i \left(\sum_{n=0}^s I_{i,j}(t - n\Delta t) w_n^i \right) \right) \right) \quad (11)$$

$$g_{k,l}(t) = h_g \left(\sum_{i,j \in N_r(k,l)} \left(h_e \left(-0,5I_{i,j}(t_1) - 0,5I_{i,j}(t_2) + I_{i,j}(t_3) \right) \right) \right) \quad (12)$$

Ahol $I_{i,j}$ egy képpont intenzitását jelenti. A fenti képletben továbbá a homogén háttér feltételezése miatt a gátlást nullának feltételezhetjük, így ezzel nem kell ebben az esetben számolni.

Tekintsük I_b háttérhez tartozó képpontértékeket és I_0 az objektumhoz tartozó képpontértékeket. A konvolúcióhoz tartozó képek közti különbséget leíró intenzitásvektor felírható, úgy, mint (I_b, I_b, I_0) az a_2 és a_1 területváltozás esetén és (I_b, I_0, I_0) az a_3 és a_2 területváltozás esetén (32. ábra).

$$g_{k,l}(t) = h_g \left((a_2 - a_1) h_e(-0,5I_b - 0,5I_0 + I_0) + (a_3 - a_2) h_e(-0,5I_b - 0,5I_b + I_0) \right) \quad (13)$$

$C = I_b - I_0 > 0$ kontrasztváltozást figyelembevéve:

$$g_{k,l}(t_3) = 0,5C(a_2 - a_1) + C(a_3 - a_2) = C(a_3 - 0,5a_2 - 0,5a_1) \quad (14)$$

Figyelembevéve, hogy a sziluettméret úgy aránylik a fókuszsíktól való távolságtól, mint a tárgyméret a tárgytávolságtól:

$$\frac{r_x}{f} = \frac{R}{d_x} \rightarrow r_x = fR \frac{1}{d_x} \quad (15)$$

Ahol a területet behelyettesítve:

$$a_x = Af^2 \frac{1}{d_x^2} \quad (16)$$

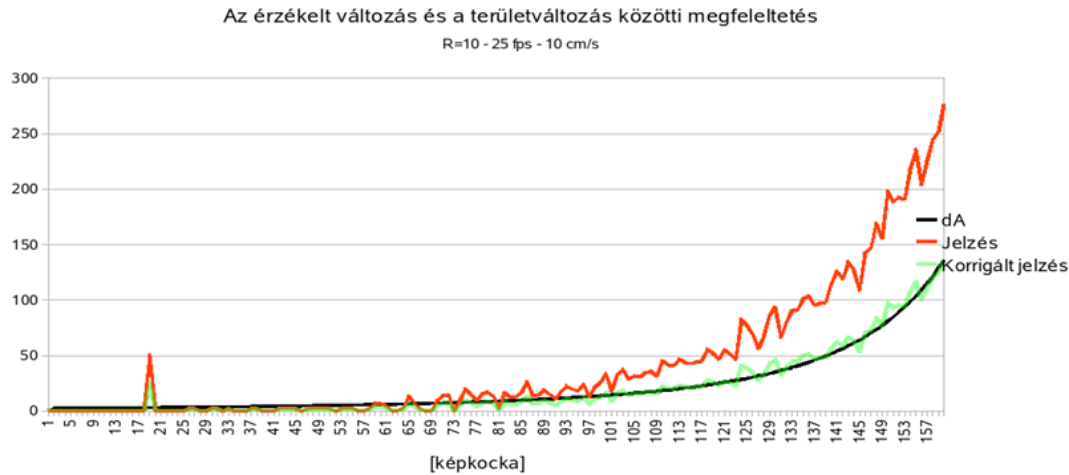
Így pedig számolható, hogy

$$g_{k,l}(t_3) = CAf^2 \left(\frac{1}{d_3^2} - \frac{1}{2(d_3 + \Delta tv)^2} - \frac{1}{2(d_3 + 2\Delta tv)^2} \right) = CAf^2 \left(\frac{1}{d_3^2} - \frac{1}{2(d_3 + s)^2} - \frac{1}{2(d_3 + 2s)^2} \right) \quad (17)$$

Ahol Δt a mintavételezés ideje, v az objektum közeledésének sebessége, míg $s = \Delta tv$ behelyettesítés az objektumtávolság két képkocka között. Az egyenlet az alábbiaként írható fel:

$$g_{k,l}(t_3) = CAf^2 \left(\frac{6d_3^3s + 21d_3^2s^2 + 24d_3s^3 + 8s^4}{2d_3^6 + 12d_3^5s + 26d_3^4s^2 + 24d_3^3s^3 + 8d_3^2s^4} \right) \quad (18)$$

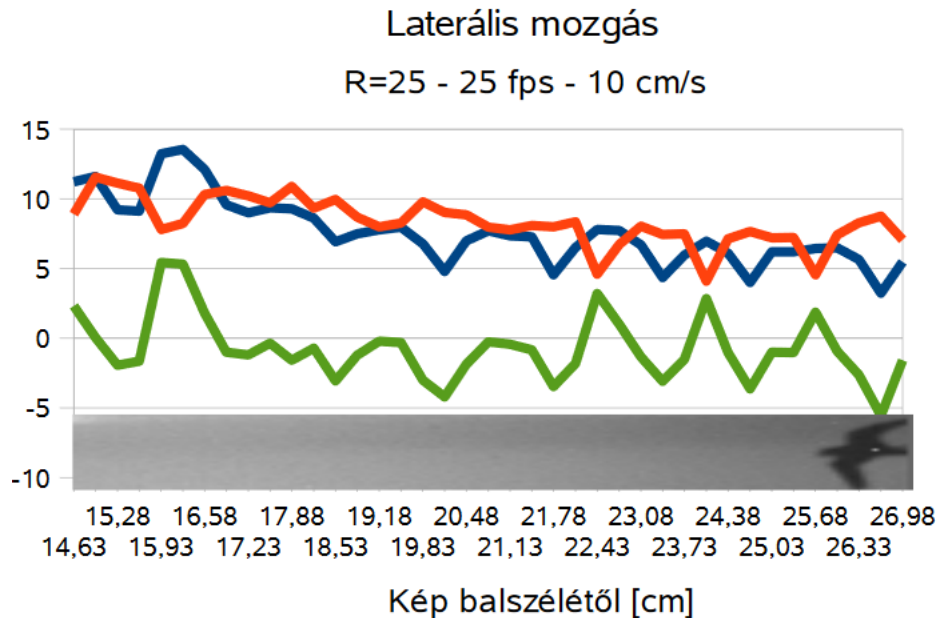
Tekintettel arra, hogy a tárgy fókuszponttól vett távolsága és a képkockák közötti elmozdulás miatt $d \gg s$ adódik, így a közeledéssel a sejtválasz $\frac{3}{x^3}$ -ként közelíthető konstans közeledés esetén, amíg a sziluettváltozás a ganglionsejt területén történik. A válasz karakterisztikáját megfigyelve látható, hogy ez a működés alapvetően utolsó pillanatos figyelemzetést ad a közeledő objektumra, mely alapvetően csak védekező reakcióra ad lehetőséget, time-to-collision számításra ebben a formában nem. Ez a működés az egér retina adottságait figyelembe véve, kevesebb, mint két másodpercet jelenthet egy egér számára egy sas közeledése esetén a menekülésre (nagy 55-85 m/s támadási sebesség, legalább 100m magasból)[101]. A 33. ábra valós Eye-RIS környezetben történt mérései összevethetőek ezzel az 1000-szeres aránnyal, a Pvalb-5 ganglionsejt kapcsán mért kb. 30-50ms-os válaszidejével kalkulálva.



33. ábra 5cm átmérőjű korong egyenletes közeledése 10cm/s sebességgel a kamera felé. A függőleges tengelyen a terület változás pixelben kifejezett $\frac{1}{x^2}$ nagysága (fekete), pirossal az erre adott $r=10$ pixel sugarú ganglionsejtválasz nagysága $[1;-0,5;-0,5]$ serkentő és $[-1;0,5;0,5]$ gátló konvolúciós ablak esetén az Eye-RIS v1.2 rendszeren, a változás receptívmezőből való kilépéséig 60cm közeledés mellett. Zölddel a jelzés x távolsággal való normálása a pixelváltozáshoz képest, figyelembevéve a konvolúciósablak $\frac{3}{x^3}$ várt változásjelzés közelítését.

A modell kalibráció kapcsán fontos megjegyezni, hogy közeledő sötét objektumokra érzékeny, tekintettel arra, hogy a természetben jellemzően világos (például az ég) háttérből közeledik a támadó. A modellben -1-szeres szorzót alkalmazva a világos objektumokra válik érzékennyé a modell, vagy az egyenirányítási lépést elhagyva és a küszöbértékeket növelve érzékennyé tehető mind sötét, mind világos objektumokra. Ez a sötét háttér előtt világos objektum közeledésére figyelmeztető út, vélhetően jelen van az egér retinában, viszont evolúciós okokból kevésbé lehet domináns. Mint láttuk a modell megfordításával ez is jelezhető.

A modell kapcsán fontos megemlíteni, hogy ha a háttér nem egyértelműen világos, hanem a szürke, vagy a közeledő objektumhoz hasonló, akkor értelemszerűen a válaszreakció is csökken, tekintettel arra, hogy kisebb lesz a kontraszt különbség, amiből a sejtválasz generálódik. Ennek megfigyelésére példa a 34. ábra, melyben a piros serkentő csatorna és a kék gátló csatorna válasza a háttér változásával együtt növekszik.



34. ábra Laterális mozgás esetén kiváltott válaszreakció kalibrált környezetben. A madár sziluettje jobbról balra egyenesen 10cm/s sebességgel mozog. A piros a serkentő, a kék a gátló csatorna számolását mutatja. Látható, hogy a világosabb háttér felé mozogva a csatornák által jelzett aktivitás párhuzamosan nő, viszont a ganglion sejten való összegzés (zöld) 0 érték körüli.

A kontrasztarányra való jelzést a képeken előzetes szegmentáció segítheti, erre vonatkozóan kezdeti kísérleteket tettem MATLAB-os környezetben, de a kutatásom idején általánosságban véve jól működő szegmentáló algoritmussal nem állt rendelkezésemre. Az akkori tapasztalatok alapján, jó szegmentáció esetén objektum szinten lehetőséget ad arra, hogy adott objektum tekintetében vizsgáljuk a sziluettméret időbeli változását.

3.2.2.4 A modell validálása, eredményei

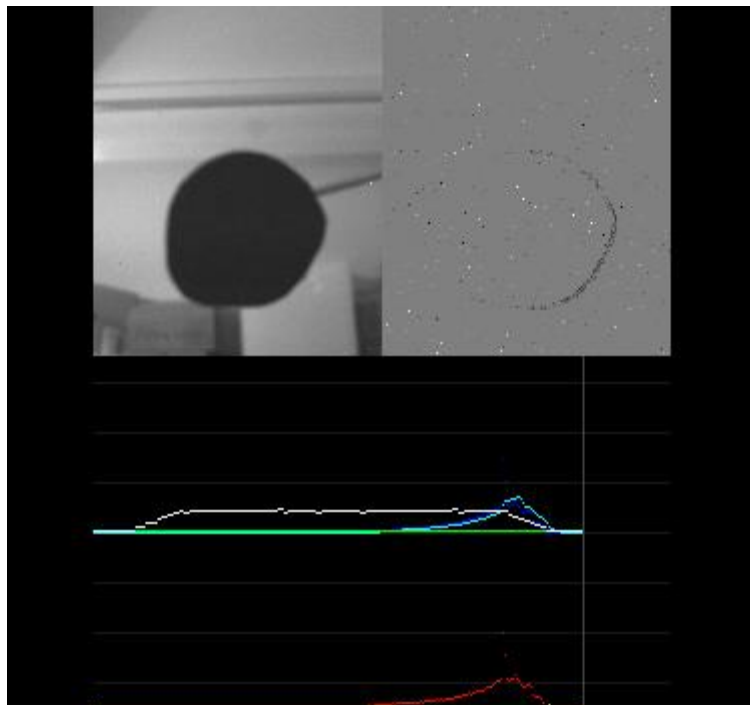
A kalibrációhoz olyan kísérleti környezetet állítottam össze, mely alkalmas volt arra, hogy a modell működése számszerűen kiértékelhető legyen. A mozgások szimulációjához egy Amtec PowerCube motorokkal szerelt CNC plotter asztalt használtam, mely egy három szabadságfokú Descartes manipulátor. A plotter asztal egy precízen vezérelhető eszköz, mely a motorok állapotáról (sebességéről, gyorsulásáról) visszajelzést képes adni. Alkalmas eszköz arra, hogy különböző mozgásminták különböző sebességek mellett megismételhetően végre lehessen hajtani és ezek a valós pozícióval összevethetőek legyenek. Az eszközzel lehetővé teszi, hogy a modell kalibrációjához valós téridőbeli stimulus mintázatokat állhassanak elő kamera számára, ismert geometriai és kinematikai információval együtt.

A felvételek az alábbi csoportokba voltak sorolhatóak:

- Különböző sebességekkel közeledő mozgás (5-10-15 cm/s),
- Különböző sebességekkel távolodó mozgás (5-10-15 cm/s),
- Különböző sebességekkel laterális mozgás (5-10-15 cm/s),
- Különböző sebességgel közeledő, de laterális irányban is elmozduló tárgy mozgás.

A plotterasztal karjára különböző formájú tárgyakat helyeztem és az Eye-RIS rendszerrel felvételsorozatot készítettem, így a modell működése és kalibrálása az Eye-RIS rendszeren és PC-s MATLAB környezetben is lehetővé vált.

Az előállt kalibrált videókon keresztül a modellt különböző paraméterekkel vizsgáltam. A különböző paramétereket a kalibrált környezetben készült felvételekkel is vizsgáltam és ezek alátámasztották az előző fejezetben bemutatott paraméterek közötti összefüggéseket (35. ábra).

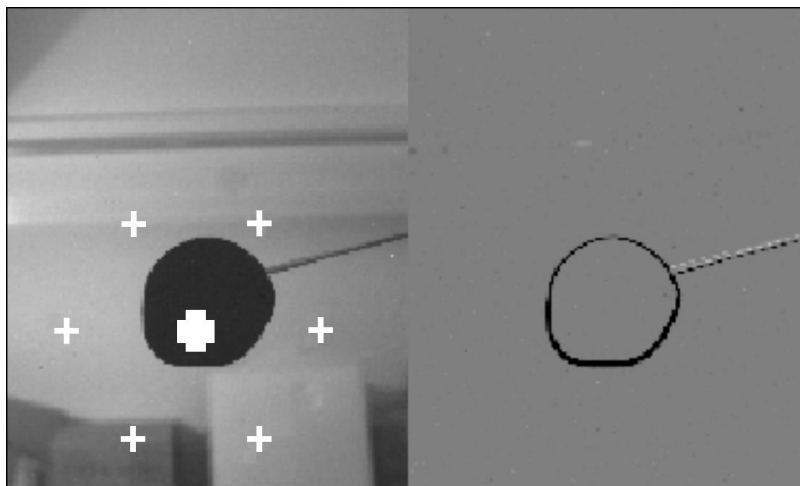


35. ábra Közeledő objektum hisztogramja 10cm/s közeledésre, majd megállásra. Fent kékkel a Pvalb-5 sejt-válasz, Fehérrel a közeledés sebessége, zölddel a laterális mozgás sebessége. Alul pirossal a serkentő sejt-válasz, sárgával a gátló sejt-válasz.

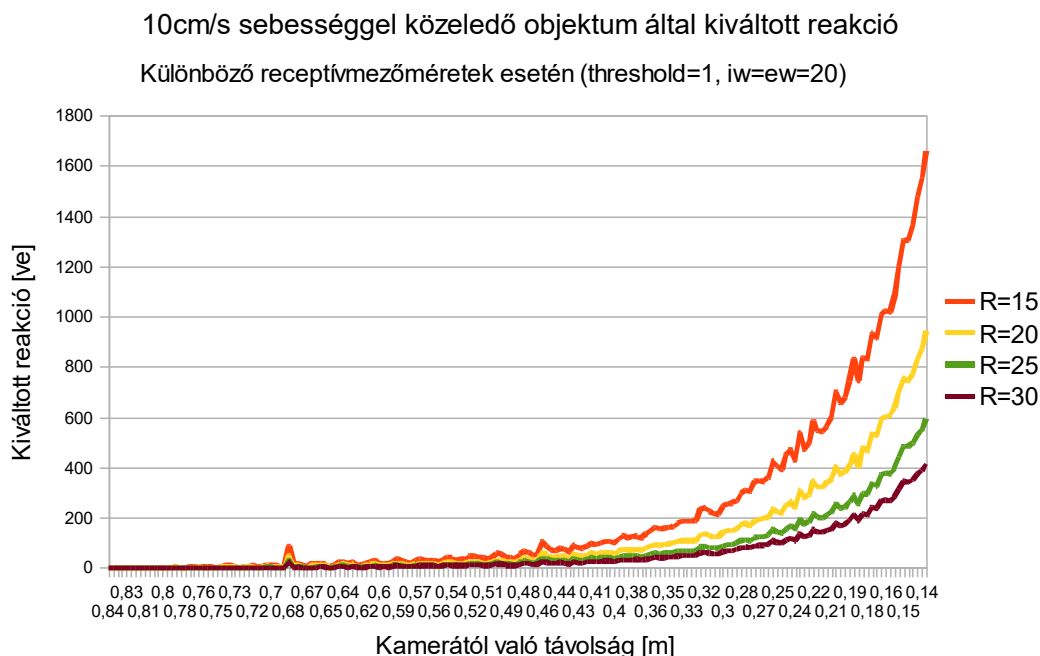
A mérések alapján a PC-s MATLAB modell és az Eye-RIS modell hasonló képpen viselkedett ugyanazokkal a beállításokkal, eltérést alapvetően az Eye-RIS analóg felépítéséből adódó számolásikülönbségek jelentették.

Az Eye-RIS modell valóban képes volt a közeledő objektumok felismerésére valós időben, ehhez kapcsolódóan élő demonstrációs környezetet is készítettem, melyen bemutatható volt a modell működés közben is.

A 36 az Eye-RIS rendszeren futása közben adott eredményt mutatja be, közeledő mozgásra. Az átfedő ganglionsejteken kiszámolja a tüzelés nagyságát, mely a közeledés közelet jelentő receptívmező centrumában a legnagyobb (a nagy fehér kereszt jelzi a mértéket a demonrendszerben). Ennek oka, hogy az adott receptívmező teljes egészében tartalmazza a közeledést, mint változást. Az átfedő részeken is keletkezik tüzelés, de jóval kisebb mértékben, ez a modell az apertúra probléma megoldása kapcsán további feldolgozást igényel.



36. ábra Eye-RIS v1.3 modell futás közben: Bal oldalon a Pvalb-5 ganglionsejt modell tüzelése a közeledő objektumra. A kereszt nagysága a ganglionsejt aktivitásának nagyságával arányos. Jobb oldalon a serkentő aktivitás látható a receptívmezők összegzése előtt.



37. ábra A Pvalb-5 sejtválasza különböző receptívmező méretek mellett

A 34. ábra zölddel jelzett ganglionsejten való összegzése megmutatja, hogy laterális mozgás esetén a ganglionsejt reakciója zéró összeg körül mozog, a háttérben tapasztalható gradiens jellegű háttérváltozás ellenére. A 37. ábra az Eye-RIS rendszeren való futás esetén a különböző receptívmező méretek melletti ganglionsejt válaszokat mutatja a közeledésre, amelyek az elvártakkal összhangban vannak.

A modell kalibrációs eredményei alátámasztották, hogy a Pvalb-5 ganglionsejt kapcsán leírt neurális működés s annak modellje képes a világos háttér előtt közeledő sötét objektumok felismerésére. A modell kalibrálható és egyszerű, hatékony implementáció adható fókuszsíkbeli érzékelő processzortömbön. A fenti modell működése és kalibrációja kapcsán releváns publikációim: [102] [103] [104] [105]

3.2.3 A Pvalb-5 alapú modell összehasonlítása más és azóta született biológiai modell eredményekkel

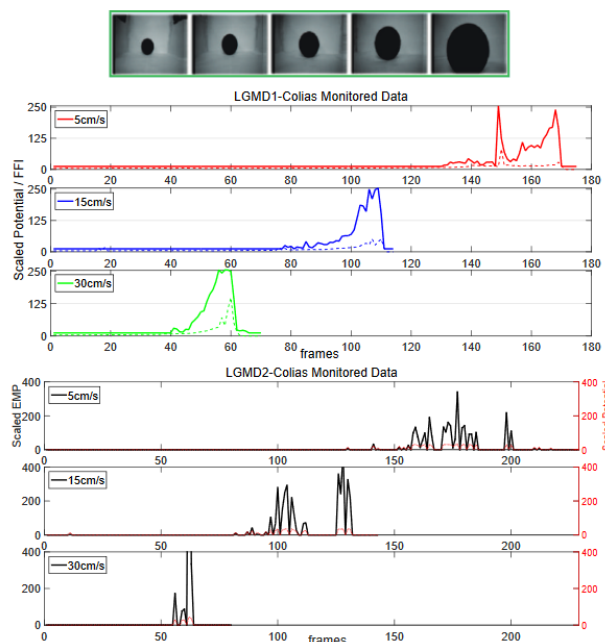
A saját, más kísérletekkel összevethető elrendezésében egy 5cm átmérőjű korong 10cm/s-al való közelítését mutatja be kb. 60cm távolságon, az Eye-RIS rendszeren, $r=10$ pixel receptívmező mellett. A rendszer válaszána időbeli lefutását a közeledésre a 33. ábra mutatja be.

Az LGMD1 és a 2016-ban modellezett LGMD2 neuron összehasonlítását a [94] cikk tárgyalja, mely hasonló kísérleti elrendezést alkalmaz: 60cm-ről egy teniszlabdát egy ferde

vájaton gurítottak le különböző sebességekkel, melyet egy 72×99 pixel nagyságú kamerával 30fpsel vettek fel. Az LGMD modellek válaszainak összehasonlító eredményeit a 38. ábra mutatja be az összehasonlító mérés egyik célja itt az LGMD2 implementáció mobil robotban való felhasználásának lehetőségeinek vizsgálata volt, annak bemutatására, hogy az LGMD1-hez képest miben másabb a karakterisztikája. A cikk megállapítása, hogy az LGMD2 neuron sebesség érzékenysége jobb az LGMD1-nél és nem utolsó pillanatos burst jellegű tüzelő választ ad, hanem a közeledés távolságával arányosabb választ.

A Pvalb-5 modellel összevetve az LGMD1 válaszgörbét az látható, hogy a Pvalb-5 10cm/s-os közeledés mellett is sokkal nagyobb tartományban volt képes (50 képkocka<) detektálni a közeledő mozgást.

A [94] megállapítása, hogy az LGMD2 az LGMD1-hez képest sokkal kisebb, de a közeledés szempontjából precízebb válaszokat képes adni. Összevetve az LGMD2-t a Pvalb-5 modellel, a válaszüidő hasonlóan alakul már képkockaszámra, viszont sokkal markánsabban képes jelezni a közeledést és a sebességet, így precízebb modellnek tekinthető. További előnye, hogy a receptívmező méretezésével az érzékenysége tovább kalibrálható (37. ábra) az LGMD modellekhez képest.



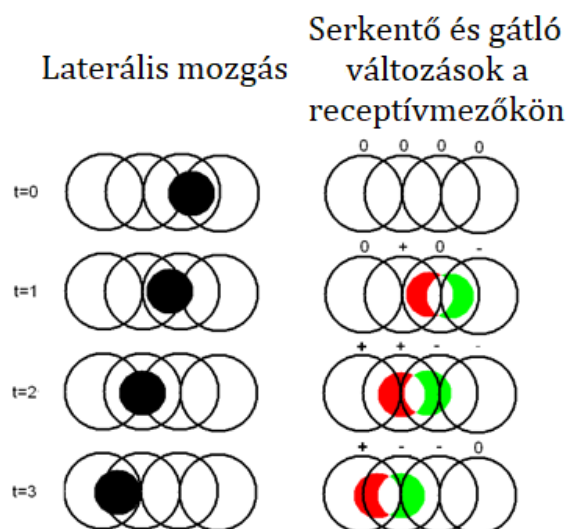
38. ábra LGMD1 és LGMD2 modellek válaszai különböző sebességű közeledő mozgásra, a modellt összehasonlító [94]-ből átvéve. Az utolsó pillanatos figyelmeztetésre alkalmas LGMD1 válaszok 0-250 közötti tüzelési értéket adnak meg 40 képkockás tartományon 5 és 15cm/s közeledési sebesség mellett, míg az LGMD2 finomabb felfutás mellett 0-100 tüzelési értékkel jelez 50 képkockán keresztül ezen az 5 és 15cm/s közötti tartományon. A Pvalb-5 ehhez képest 0-275 közötti tüzeléssel reagál 100 képkockán át 33. ábra

3.3 Közeledés irányának meghatározása Pvalb-5 sejt működési modelljének felhasználásával

A Pvalb-5 ganglionsejt harmadik karakterisztikus paramétereinek tárgyalása során már jeleztem, hogy az átfedő receptívmezők vizsgálata további vizsgálatra érdemes, valamint utalást tettem arra is, hogy a sziluett receptívmezőbe való belépésekor, vagy onnan való kilépéskor erős serkentés, illetve erős gátlás történik. Ez a működés lehetőséget teremt további összefüggések kutatására a modell kapcsán, így annak vizsgálatára, hogy a közeledő objektum viselkedése kapcsán milyen további válasz adható a modellt felhasználva, valamint, hogy az apertúra probléma megoldására a modell milyen lehetőségeket ad több ganglionsejt sejtválaszát együttesen felhasználva.

3.3.1 Az eredeti modell módosítása közeledés iránykomponensének számításához

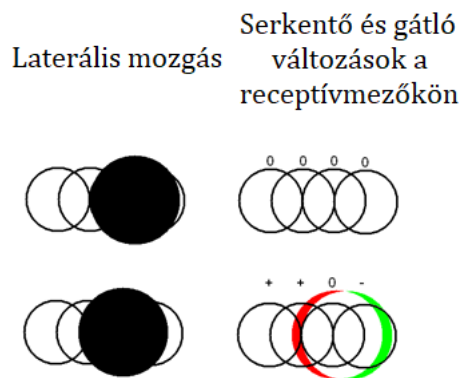
Közeledéskor az összes érintett mezőben válasznövekedés tapasztalható, a legnagyobb a tárgy mögött található receptívmezőben, amíg annak méretét eléri. Abban a pillanatban, ahogy a növekedés kikerült a receptívmező területéről, a tüzelés abbamarad, viszont az átfedő receptívmezők még továbbra is működésben vannak, s szumma válasz értékük tovább növekedik. Ha az objektum távolodik, akkor a gátló csatornák lesznek aktívak és a működés pontosan ellentettje a közeledéskor tapasztalhatónak.



39. ábra Laterális mozgás közben a serkentés és a gátlás térben egyidőben teljesen eltérően aktiválhatja a ganglionsejteket.

A közeledő mozgás érzékelés modellnél ismerttettem, hogy laterális mozgás esetén a receptívmezőn belül a serkentés és a gátlás jelzés kiegyenlítődik. Átfedő receptívmezőknél, az objektum például a képsíkon történő balra haladása esetén az egyes átfedő receptívmezőkre hat egyaránt gátlás és serkentés is, mint ahogy azt a 39. ábra szemlélteti. A $t=0$ időpontot követően a serkentő csatorna válasz („+”) és gátló csatorna válasz („-”) akkor jelentkezik, amikor az objektum sziluett folyamatosan belép (pirossal jelzett változásérték), vagy folyamatosan elhagyja az adott receptívmezőt (zölddel jelzett változásérték). A receptívmezőn belüli mozgás esetén a kettő a vártnak megfelelően kioltja egymást.

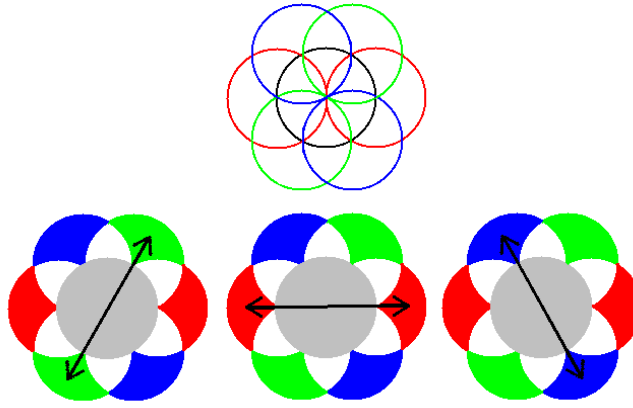
Abban az esetben, ha az objektum mérete nagyobb egy receptívmezőnél, akkor az alábbi ábrán látható módon alakulhatnak ki potenciálkülönbségek. Itt az látható, hasonlóan viselkedik, mint az előző, receptívmezőnél kisebb sziluett esetén, viszont egyes ganglionsejteket esetleg nem stimulál. Ekkor a reakció a sejt környezetében jelenik meg. Mindkét viselkedésnél megfigyelhető, hogy a változás itt is potenciálkülönbséget okoz az egyes ganglionsejt-válaszok között.



40. ábra A receptívmezőnél nagyobb objektum laterális mozgása esetén kialakuló csatornaválaszok.

3.3.2 A matematikai modell bemutatása

Az eredeti modellt úgy módosítottam, hogy a ganglionsejt összegzésénél az utolsó egyenirányító lépést kihagytam, a küszöbérték számítás egy szignum függvényvel vett szorzat. Ezzel az előzőekben bemutatott viselkedést megfigyelve és kihasználva a ganglionsejtre már kiszámolt potenciálértékeket, egy adott ganglionsejtre fel lehet írni az aktivitáskülönbségeket általánosságban is az átfedő receptívmezőkre, mint amelyet példaként a 28. ábra mutat be.



41. ábra Demonstrációként, négyszeresen átfedő receptívmező (28. ábra) felépítésében a nyíllal jelzettekkel felírhatók potenciálkülönbségek.

Egyenletesen fedő rétegeket feltételezve a 41. ábra szerint számolhatóak egymás melletti receptívmezőkön megjelenő potenciálkülönbségek az átfedő (szürke) receptívmezőre. Az egymás melletti mezők egymáshoz képest lévő középpont elhelyezkedése megadja egy vizsgált mozgásirány iránykomponensét. A mezőkön a sziluett elmozdulásának nagysága pedig megadja, hogy az adott irányon ez a mozgás mennyire jelentős.

Több nem átfedő szomszédságot vizsgálva és számolva az eredők megadják az elmozdulás irányát és nagyságát, vagyis a lokalitást több szomszédos receptívmező középpontjára számolva egy négyes átfedő hálózaton az alábbiak szerint adható irányvektor az elmozdulásra a fenti ábrán látható átfedéseket figyelembe véve az alábbi lineáris kombináció számolásával:

$$\vec{v}_e = \sum_{i \in n} a_i \vec{v}_i \quad (19)$$

ahol:

- a_i a potenciálkülönbség nagysága és előjeles iránya két szomszéd között,
- \vec{v}_i két egymásmelletti receptívmező középpontjai közötti normálvektor

Ez a 41. ábra esetében látható átfedés esetén a következő képpen számolható:

$$\begin{pmatrix} x \\ y \end{pmatrix}_{\text{középső}} = \begin{pmatrix} (rR - rL) + \sqrt{(bR - bL) + \sqrt{(gR - gL)}} \\ \sqrt{(bR - bL) + \sqrt{(gR - gL)}} \end{pmatrix} \quad (20)$$

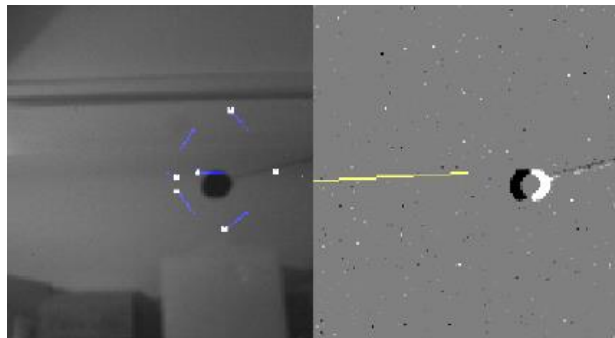
Ahol r : piros, g : zöld, b : kék receptívmezők tüzelési értékei a középpontokhoz képest balra (L), vagy jobbra (R). Ha a receptívmező elegendően kicsi, akkor ezek a vektorok a lokális változások számolására alkalmasak, így például régiók optikai áramlásának számolá-

sára. Közeledő mozgás esetén az értékek számolásánál ugyanúgy kialakulnak potenciálkülönbségek, melyek a mozgás irányába mutatnak az epipoláris geometriai térben, vagyis ezzel a közeledő objektum mozgásának irányultsága is megadható. Ez lehetőséget ad egy adott ganglionsejt tekintetében az apertúra probléma részleges megoldására is, mivel így a rendszer lokális válaszaival számolni lehet.

3.3.3 A modell implementálása és validációja

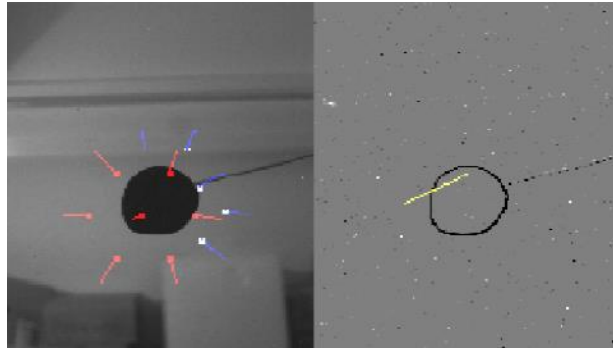
A felírt potenciálkülönbségen alapuló modellt PC-s MATLAB környezetben vizsgáltam kalibrált környezetben és általánosságban. A kalibrált környezetben történő vizsgálatokhoz a retinacsatorna modell vizsgálatához felhasznált felvételeket vettem alapul.

A modell vizsgálatához egy 5 cm átmérőjű kör alakú lemezt és négy átfedő receptívmezőt használtam. A laterális mozgások vizsgálata során különböző sebességekkel vizsgáltam a rendszer választ. A kapott válaszok alapján a modell az elvárt módon viselkedett (42. ábra), így kisebb sebességeknél kisebb irányvektorral is jól mutatta meg a mozgás irányát. A válasz nagysága és iránya összhangban volt az elvártnal.



42. ábra Laterális mozgás esetén a bal oldali képen látható iránykomponensek az aktivitás irányát adják meg. Jobb oldalon látható a gátló és serkentő csatorna aktivitás mintázata és a potenciálkülönbségek által számolt mozgásirány, mely felé a sárga, mozgás irányát jelző mutató is áll. A változások a 9.1 függelékben lépésenként is szerepelnek.

A közeledő mozgásokról készült felvételnél felfedezhető laterális irányú komponens, mely azért vált láthatóvá, mert a tárgy nem tökéletesen a fókuszsíkra merőlegesen, a fókuszpont felé mozgott. A közeledés irányultsága szintén megfelelt az elvártnak, ehhez kapcsolódó szimulációs eredményt mutat be a 43. ábra.



43. ábra Közeledő, mozgás esetén a bal oldali képen látható iránykomponensek szerinti aktivitás irányt adtak, a jobb oldalon látható, hogy csak a serkentő csatorna ad választ. A jobb oldalon sárgával jelölt kis iránymutatás észrevette, hogy van egy enyhe laterális irányú mozgása is az objektumnak.

3.3.4 A modell működése valós környezetben

A modell működését vizsgáltam valós környezetben is: fix kamerapozíciót választva a modell megfelelően felismerte a képen történő laterális és közeledő mozgásokat. Mozgó platformon – autóra felszerelve a kamerát – vizsgálva a modellt először a kép stabilizálására volt szükség, mivel egy-egy képkocka között is nagy elmozdulások történhettek. A képet stabilizálva viszont távoli tárgyak esetében (pl. horizont vonalának környezetében található közeledő fák) a választott receptívmező méret és a küszöbértékek függvényében képes volt értékelhető választ adni a modell a közeledés irányára vonatkozóan sötét tárgyak, világos háttér előtt való mozgásakor.

A modell működését befolyásolták a lencsén felcsillanó erős fényhatások (nap), valamint a nem konstans fényességű területek, így a valós környezetben történő alkalmazása önmagában nem volt értékelhető. Egyszerűbb környezetben a modell a megfogalmazottak szerint képes volt jelezni az irányt, így a modell alkalmazásának előfeltétele az előzetes képszegmentálás, melyben objektumonként nézve lehet következtetéseket levonni.

A modell bemutatása kapcsán publikációm: [104]

3.4 A biológiai alapú modell lehetőségei és korlátai

A Pvalb-5 ganglionsejt modellezése alkalmas volt arra, hogy igazolja a neurobiológusok által talált retinacsatorna működését. Az analóg processzortömbökön való implementáció valós idejű szimulációk számára ad lehetőséget arra, hogy a felírt modellek akár közvetlenül is összehasonlíthatók legyenek a valós biológiai rendszerekkel. Másrészt ha áttörést sikerül elérni az analóg érzékelő processzortömböknél fennálló korlátoknál (zaj, méretkorlát [106]), akkor lehetővé válhat széleskörű elterjedésük is.

A biológiai modellek implementációja mindemellett a képfeldolgozási algoritmusokkal foglalkozók számára jelenthet megoldási alternatívát, egy-egy algoritmikusan nehezebben feldolgozható problémakör megoldására. Tipikusan ilyenek a képfolyamokon megtalálható események felismerései, mint amilyen a közeledés is.

A biológiai alapú algoritmusok terjedését viszont mindenképpen gátolja, hogy a normál algoritmusokhoz képest más felépítésűek, valamint számításintenzívek. Ezen feladatok megoldására az analóg, CNN felépítésű chipek praktikusán és jól használhatók lennének, viszont ezen analóg chipek jelenlegi fejlettsége nem teszi lehetővé nagy képméretűek kezelését a nagy analóg zaj miatt. Érdemi képfeldolgozási műveletet ezen architektúrákon még nem lehet elvárni, így egyelőre marad a szimulálásuk lehetősége.

3.4.1 A Pvalb-5 alapú modell korlátai

A közeledő mozgásra érzékeny modell első karakterisztikus paramétere, mint láttuk a konvolúciós ablak meghatározása. Ezen paraméter megválasztása a képsebesség függvényében alapvetően határozza meg a rendszer viselkedését valós környezetben való alkalmazáskor. A felvétel készítése során a nagy változások képesek hibás közeledés detekciót eredményezni. Ilyen nagy és befolyásoló jellegű változás az objektív, vagy valamilyen tárgy megcsillanása, a textúra gyors változásai, a kamerában működő esetleges tömörítési algoritmusok artifakt jelenségei, vagy azok hibája. Ezeket a változásokat akár nagy, közeledő jellegű változásnak érzékelheti a modell, ha kicsi a konvolúciós ablak mérete, mivel a változás az alak méretéhez képest jelentős.

A modellnek mindenképpen komoly korlátja az OFF típusú működés, mely invertálva ON-ra is alkalmazható, mindemellett a kis kontrasztkülönbség esetén kevésbé reagálóképes. Ezt a hátrányt a kép előzetes szegmentálásával el lehet kerülni.

A közeledés irányultság meghatározásának valós felvételeken történt vizsgálata során a valós környezetben és bonyolultnak tekinthető felvételeken sokkal inkább optikai áramlásra jelzett a modell közeledés irányt, valamint autós felvételeken, a periodikusan változó intenzitású LED-es fényforrások a receptív mező modell számolását befolyásolta.

Ez utóbbi jelenséget annyira figyelemreméltónak ítéltém, hogy elkezdtem ennek vizsgálatával is foglalkozni és a modellt tovább gondolni, miszerint impulzusjelenségek, így impulzusüzemű fényforrások felismerhetők lehetnek a ganglionsejt modelljének kisebb változtatásával.

4 Periodikusan változó fényintenzitású LED-es jelzőlámpa felismerése

A bevezető fejezetben a változások felismerésének problémakörét érintettem már. Egy videófelvételen a változás észlelése a kontextustól is függ, így a konkrét eseménydetektáláson túl (például eltűnik, megjelenik, elmozdul, megváltozik valami, stb.) befolyással van a történések elemzésére, következtetések levonására, vagyis a videófelvétel magasszintű, szemantikai értelmezésére.

Az objektumok minőségváltozásával kapcsolatban is beszélhetünk változás felismerésről és annak értelmezéséről. Egy-egy jellemzőt felhasználva további szemantikai következtetés vonható le, mely szintén kurrens téma [107].

A fényt adó objektumok felismerése, elhelyezkedésük meghatározására a képfeldolgozás szemantikai értelmezésében fontos szerepet játszik: egy lámpa fel van-e kapcsolva, vagy nincs, valamint a virtuális valóság alkalmazások esetében konkrétan felhasználható az az információ, ha ismert az egyes fényforrások helye a térben. A fényforrások felismerését főként a képet statisztikai megközelítésű módszereket felhasználó algoritmusokkal elemzik, melyek a fény viszonyokra és az eloszlásukra építenek, akár több fényforrás egyidejű megvilágítása esetén is. [108] A fényforrások felismerése viszont a képfeldolgozás alacsonyabb szintjein, például triviálisan a digitális fényképezés lokális expozíciós idejének állítása során is fontos szerepet játszik.

Napjainkban az ADAS rendszerek számára is fontos jelentőséggel bír az aktív fényforrások felismerése, mivel az út mellett ezek jellemzően valamilyen információt hordozhatnak és így gyorsabb felismerésük támogatásával a közlekedés biztonságát képesek javítani: így a közlekedési lámpák felismerésével, jelzésekük meghatározásával, más járművek fényeinek észlelésével, vagy akár világító táblák értelmezésével. A világító táblák jellemzően kritikus helyeken vannak kihelyezve, a jelzőlámpák jelzésekének helyes felismerése pedig már biztonságkritikus feladatnak tekinthető, főként az önvezető autók terjedése miatt. Bár a jelzéseképeket elvileg rádióhullámok segítségével lehetne közölni, de erre mind a mai napig nincsen általános és bevett szabvány, valamint nem látszik olyan jogszabályi környezet változás sem, mely a közeljövőben ezt kötelezővé tenné.

A jelzőlámpa felismerését nehezítheti a Nap állása, elhelyezkedése, mivel ekkor a kamerákban lévő expozíciós időt befolyásoló algoritmusok a rekeszidőt befolyásolják, így pedig a felismerés nehezebb.

A világon mind elterjedtebbek a LED-es jelzőlámpák, melyek jelzésiképe még rossz fényviszonyok mellett is jobban felismerhető a hagyományos izzósokhoz képest. Ennek több oka is van: részint a látható fény hullámhossza szűkebb és nagyobb lehet (150lm/W), másrészt működésük jellemzően a retinát is jobban befolyásolja: észrevehetőbbek ellenfényben is, mivel kihasználja a retina változásra való érzékenységét, mely ezen lámpák működéséből is adódik. A LED-es autós közlekedési lámpákat impulzusüzemben működtetik a hálózati feszültség egyenirányítása és egyszerűbb áramköri felépítése miatt, valamint így a diódák hasznos élettartama is növekszik. Ez az impulzus-szerű működés viszont szintén hatással lehet a jelzőlámpafelismerő algoritmusokra, mivel a lámpa fényessége nem konstans, így nem megfelelő sebességgel készített felvétel esetén, a lámpa impulzus-szerűen vibrál (44. ábra). Ennek oka, hogy a videókészítés során a képkészítés sebessége alacsonyabb, mint a fényforrás frekvenciája.



44. ábra 30 fps sebességgel készült felvétel egy piros lámpáról.

A Pvalb-5 ganglionsejt modelljénél bemutatott egyik mellékhatás az impulzusszerű nagy változásokra való reakcióképesség volt, melyet a jelzőlámpát tartalmazó képeknél észleltem. Az alapfeltevésem az volt, hogy receptívmezőt megfelelő méretűnek beállítva, valamint a konvolúciós ablakot is megfelelően megválasztva, a receptívmezőn számolt gátló és serkentő csatorna értékét számolva egy periodikus értékváltozás kapható. A modell ezt a már bevált három hosszú ablak mellett egyértelműen kimutatta, viszont gyakorlati alkalmazása a felismerési feladatokra nem volt alkalmazható, mivel nem tudta szelektíven kezelni a vibrálás jelenségét a környezet más jelenségeihez képest. Mindezek ellenére a jelzőlámpa jelzésikép felismerés témakörére, ezen időbeli változás alapú felismerésére helyeződött át. Az alábbiakban bemutatom, hogy egy általános kamerát felhasználva lehetséges alul-mintavételezés frekvencia mellett jelzésikép felismerése.

4.1 Szín alapú szegmentálás

Az érdekes régiók kijelölése alapvető fontosságú, hiszen ez az alapja egy-egy objektum minőségi változásának a megfigyelésére. Számos technika létezik rá, mint arról a korábbi fejezetekben szó volt. Részint a LED-ek esetében a színszegmentáció tipikusan alkalmas

lehet erre, mivel a fényhullámhosszának a szórása nagyon csekély, vagyis alapvetően jól szegmentálható és gyors módszer, mivel megfelelő szintérben ez egy küszöbérték meghatározását jelenti a szórás környékén. Hagyományos lámpák esetén is ez bevett lépés, különböző szinterek [109] [110]. A jelzések színe kapcsán szabványokat kell a gyártóknak betartani (MSZ EN 50556 és MSZ EN 12368 harmonizált szabványok). A jelzések felépítése országonként különbözhet, viszont alapvető jellemzőik nemzetközi szabványoknak is meg kell feleljenek. A vasúti szabványok ezektől biztonságkritikus kialakításuk miatt több ponton eltérnek.

4.2 Alul-mintavételezett frekvencia elemzés

Budapesten is alkalmazott jelzőlámpák kapcsán konkrét gyártói információ nem áll rendelkezésre a működési frekvencia kapcsán, viszont videófelvételes méréseim alapján 100Hz-nek adódott, melyet más tanulmány is megerősít [111]. Ez az 50 Hz-es hálózati frekvencia egyenirányításából adódik.

30 Hz-el készült videófelvétel készítésekor látható a 100Hz-es folytonosan változó jel vibrálása, azonban Shannon mintavételi tétele [112] alapján ezt a folytonos idejű jelet legalább kétszeres mintavételi frekvenciával kellene készíteni ahhoz, hogy az eredeti frekvenciaspektrum előállhasson és alkalmazhatóak legyenek frekvenciafelismerési eljárások. Tekintettel arra, hogy alul-mintavételezés (aliasing) esetén átlapolódik, eltolódik a frekvenciaösszetevő [113], így vibrálásként látható ez a felvételeken.

A látszólagos frekvencia (AF) értéke megegyezik a mintavételi frekvencia legközelebbi egészszámú többszöröse (CIMS F) és a bemeneti jelalak frekvenciája (IF) közti különbség abszolút értékével:

$$AF = |CIMS F - IF| \quad (21)$$

Számos jelfeldolgozási terület van azonban, amelynél az aliasing mellett kell jeleket feldolgozni, így ekkor olyan eljárást kell alkalmazni, ami képes lehet becsülni a frekvencia-komponenseket, az alul-mintavételezésre jellemző jellemzőkből [114].

4.2.1 Goertzel algoritmus

A tipikusan valamely frekvenciakomponens felismerésére a Fourier transzformációt (DFT, FFT algoritmusok) szokás felhasználni, amelyek minden frekvenciakomponensre

kiszámolják az egyes frekvenciaösszetevők nagyságát. Ezek egyszerű, jól párhuzamosítható algoritmusok, viszont problémájuk, hogy az Euler féle formulát használják, sok frekvenciakomponensen végig számolva.

Nagysebességű kamerás megoldásoknál készült a LED-es lámpák felismerésére ezen algoritmus háttérrel, viszont számításuk közismerten költséges [111]. Emiatt az olyan analóg processzortömbön, mint amilyen az Eye-RIS is, nem használhatóak, csak a rendszerben lévő FPGA-n.

```
dft(inreal[], inimag[], outreal[], outimag[])
{
    N = length(inreal);
    for (k=0; k<N; k++)
    {
        real = 0;
        imag = 0;
        for (n=0; n < N; i++)
        {
            angle = (2*PI*n*k)/N;
            real = real+(inreal[n]*Cos(angle)+inimag[n]*Sin(angle));
            imag = imag+(-inreal[n]*Sin(angle)+inimag[n]*Cos(angle));
        }
        outreal[k] = real;
        outimag[k] = imag;
    }
}
```

([80] alapján)

$$X_k = \sum_{n=0}^{N-1} x_n \cdot e^{-\frac{i2\pi}{N}kn} = \sum_{n=0}^{N-1} x_n \cdot \left[\cos\left(\frac{2\pi kn}{N}\right) - i \cdot \sin\left(\frac{2\pi kn}{N}\right) \right] \quad (22)$$

Olyan eljárást kellett keresnem, amely alkalmazható lehet az Eye-RIS rendszeren is, vagyis az Euler féle formula exponenciális számolását elhagyja, vagy egyszerűsíti, valamint szempont volt, hogy jól párhuzamosítható legyen és megengedhető, hogy ne a teljes frekvenciaspektrumon legyen végig számolva, hanem csak néhányon.

Ennek a kritériumnak a Goertzel [115] algoritmus felelt meg, amelyet tipikusan ismert frekvenciák felismerésére használnak (pl. kifejezetten a tone üzemmódú telefonok által gerjesztett frekvenciaösszetevőinek felismerésére). Tekintettel a problémakörre, miszerint ismert a mintavételezési frekvencia, valamint ismert a jelalak is, így ez az algoritmus a felhasználási kritériumoknak megfelelt. Az algoritmus az Euler formulát ez is használja, viszont ez ebben az esetben előre számolható, illetve a szinuszos és koszinuszos összetevőkkel a kívánt frekvenciákra és konstansként kezelhető. További jó tulajdonsága, hogy egyszerűen és gyorsan számolható kizárólag keresett frekvenciákra az alábbiak szerint, ahol $x(n)$ a bemenő jel:

$$S_k(n) = x(n) + 2 \cos\left(\frac{2\pi k}{N}\right) S_k(n-1) - S_k(n-2) \quad (23)$$

$$Y_k(n) = S_k(n) - W_N^k S_k(n-1) \quad (24)$$

```

realW = 2*cos(2*pi*k/N);
imagW = Sin(2*pi*k/N);

S1 = 0;
S2 = 0;
S = 0;
for (n=0; n<N; ++n)
{
    S = x(n)+(realW*S1)-S2;
    S2 = S1;
    S1 = S;
}
outreal = 0.5*realW*S1-S2;
outimag = imagW*S1;

```

[80] alapján

A fenti algoritmus frekvenciánként az Eye-RIS-ban definiált képmemóriával gyorsan számolható, a fenti operátorok rendelkezésre állnak hozzá, csupán a szinusz és koszinusz számolása esetben szükséges az értékek előzetes előállítására FPGA-n a gyors számoláshoz.

```

realW = 2*cos(2*pi*k/N); //FPGA művelet cos miatt
imagW = Sin(2*pi*k/N); //FPGA művelet sin miatt

LAM_1 = 0; //CNN művelet
LAM_2 = 0; //CNN művelet
LAM_0 = 0; //CNN művelet
for (n=0; n<N; ++n) //ciklus FPGA közreműködéssel
{
    LAM_0 = LAM_x(n)+(realW*LAM_1)-LAM_2; //CNN művelet
    LAM_2 = LAM_1; //CNN művelet
    LAM_1 = LAM_0; //CNN művelet
}
outreal = 0.5*realW * LAM_1 - LAM_2; //CNN művelet
outimag = imagW * LAM_1; //CNN művelet

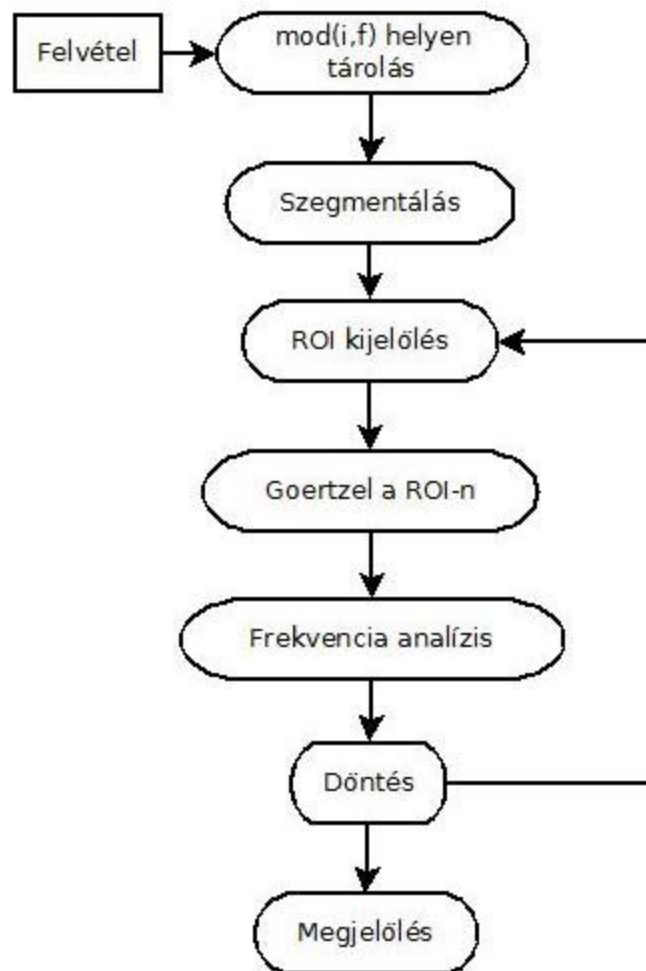
```

A fenti pszeudokódban a ciklusban található műveletek a LAM analóg memóriacellák közötti műveletek, a ciklus hívások pedig az FPGA-ról kell kezdeményezhetőek, ahogy a szinusz és koszinusz számolás is. Az Eye-RIS v1.3 rendszeren a fenti algoritmus viszont nem implementálható, mivel az analóg processzortömb 6 analóg memóriacellája nem elegendő a korábbi képek tárolásának szükséglete miatt a számoláshoz. A ciklusban szereplő LAM_x hívás során az FPGA digitális memóriához hozzáféréssel megvalósítható, viszont időigénye viszonylag nagy a rendszerben. Az analóg memória hiánya miatt így az elvi előnye az implementációnak emiatt elveszik, a valós idejű futtatás nem megoldható ezen írási-olvasási sebesség miatt.

4.3 LED jelzőlámpa felismerése Goertzel algoritmus segítségével

A Goertzel algoritmus alapú felismerés algoritmusát mutatja be a 45. ábra. A felvételt követően egy körkörös bejárat tömbbe kerülnek az egyes képkockák. A képkockákon szín alapú szegmentálás történik, majd a tulajdonságok alapján (pl. KKN alapján az előző képkockához való viszonya) az egyes régiók felcímkézésre kerülnek (ha korábban címkézettek voltak, akkor az pedig felülvizsgálatra kerül).

Az egyes régiókon ezt követően végrehajtódik a frekvenciaelemzés, melynél a csúcs alapján döntés történik.



45. ábra A jelzőlámpa kereső algoritmus egyszerű feldolgozási lánc, ahol a felvétel során mindig az utolsó f képkockát tároljuk el (a helykiosztás modulo alapú).

Az algoritmust először a Pvalb-5 receptívmezőkön megfigyelve azt tapasztaltam, hogy álló képen alkalmas volt a képrészleten frekvencia felismerésre, színszűrés nélkül, viszont folyamatosan problémát jelentettek a zaj jellegű hatások, főként a szél által gerjesztett, becsillanó jellegű képi események. Ezek részint ugyanúgy jelezhatték egy adott frekvencia felharmonikus jelenlétét, de erősen zajszerűen viselkedtek. Ezeket a modellben nem tudtam kiküszöbölni, így a receptívmező alapú felismerést itt elvettem.

A színszegmentációval viszont jól megtalálhatóak voltak a lámpák, akár néhány pixeles méretben is. Egymás melletti frekvenciakomponensek számolásánál láthatóvá vált, hogy a LED esetén, az elvárásoknak megfelelően, a szomszédos frekvenciakomponensek számottevően gyengébbek voltak, így a módszert kiegészítettem a szomszédos frekvenciakomponensek számolásával is, hogy a zaj jellegű felismerés kiszűrhető legyen.

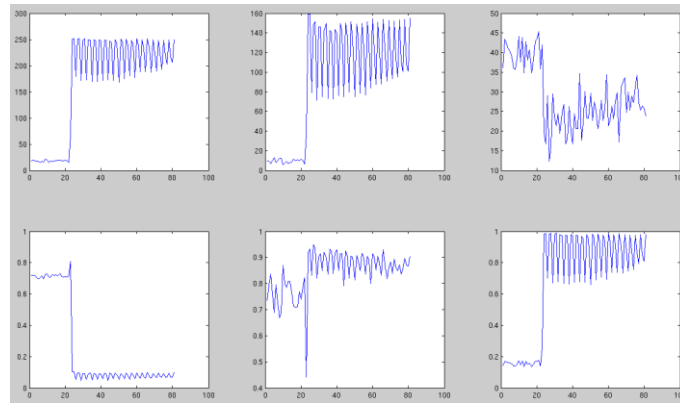
Hatással volt a modell felismerésére még a bemeneti képfolyam memóriája. Ez a felismerési kívánt frekvencia kapcsán számolható, ami alapján a 100Hz-es jel észleléshez 30fps mintavételezés esetén legalább 11 mintával való számolás szükséges.

Ezt követően a modell álló helyzetben képes volt felismerni a képfolyamon található meghatározott frekvenciakomponenseket, viszont mozgó platformon viszont csak tracking algoritmus használata esetén, az optikai áramlás miatt.

Bár a néhány pixel nagyságú távoli lámpákat a szegmentálás előhozta, viszont a JPEG tömörítés artifakt hatása a frekvenciaelemzésre hatással volt, sokszor elrontotta a frekvenciaelemzést.

Az álló platformos kísérletek azt mutatták, hogy az algoritmus jó ROI kijelölés mellett használható lehet képfolyamon lévő periodicitás felismerésére. Ehhez kapcsolódóan a ROI kijelölést az impulzus üzemben működő LED világítótestekre szűkítettem. Kísérleti mérések alapján megállapítottam az alábbiakat:

1. RGB színtérben az egyes összetevők felismerésénél mindhárom tartomány egyszerre változik, mivel együtt adják a világosság értékét.
2. HSV színtartományban a Színesség (Hue) értékének 2-3%-os megváltozása mellett a Szaturáció értéke 5-10%-al tér el a maximálistól, az Érték (Value) pedig 20-40%-al is.



46. ábra Lámpa sárgára váltás komponensei. Felső sor: RGB komponensek, Alsó sor: HSV komponensek



47. ábra Álló helyzetben felismert jelzéskép. A legkisebb felismert méret $r=2$ pixeles kör, de a JPEG artifact jelenség miatt az algoritmusnak ez a legtöbb esetben azért problémát okoz.

A 2. számú megállapítás azt mutatta, hogy a Hue értéke alkalmas ROI kijelölésre a Szaturációval együtt, míg a frekvenciafelismerés kapcsán a Value érték alkalmas erre.

Mozgó platformos felismerés kapcsán, a folyamatos mozgás ellenére, még viszonylag nehezebb környezetben is alkalmazható volt a módszer (egyszerűbb városi környezet, ellenfény, optical flow jelensége, mozgó fák által keltett vizuális zajok, ablakok megcsillanása). A Hue-t és Szaturációt felhasználó szegmentálás a LED-es lámpákat képkockáról, képkockára képes volt megtalálni. Szegmentációs hibát a kamera integrálási idejének környezethez való folyamatos adaptálódása, valamint a kamerában jelenlévő tömörítési eljárás miatt a színek torzulása és a JPEG artifact hiba.

A felismerési hibákat egy legközelebbi szomszéd (KNN) alapú tracking algoritmussal [116] és a köztes pontok becslésével csökkentettem.

4.4 Eredmények

Újpalotán városi környezetben, egy soklámpás kereszteződés környezetében, ahol a panelház ablakok csillogása és az őszi falevelek miatt a környezet már nehéznek tekinthető, egy 13 000 HD képkockát tartalmazó felvételsorozatot készítettem tiszta időjárásban. A felvételen képkockánként kézzel megjelöltem az egyértelműen 3-4 pixelen is már látható lámpák jelzéseképét, a jelzésekép kiterjedését. Az így összeállt referenciaadatokkal hasonlítottam össze a megoldást.

A 13 000 képkockán 5 399 helyen volt piros jelzésekép. A piros jelzéseképeket 86,76%-ban ismerte fel az eljárásom és 8% körül alakult a hibás pozitív (fp) eredmények száma ebben az egyszerű modellben. Az alábbi táblázat megadja az eredményeket:

	Felismert Pozitív	Felismert Negatív
Pozitív	4693(tp)	716(fn)
Negatív	1400(fp)	17304(tn)

tp: helyesen talált, fp: helytelen találat,
tn: helyesen nem talált, fn: nem talált meg létezőt.

Ez a piros jelzésekép esetében az alábbi eredményeket jelenti az algoritmus minőségét tekintve a fenti táblázat alapján:

- Precision= 0,77
- Recall= 0,87

A sárga, de különösen a zöld esetében rosszabb eredmények születtek, melyek a kamera integrálási idejének változásából, valamint a fehéregyensúly állításából adódtak. A zöld esetében egyértelműen megfigyelhető volt, hogy ugyanazon látott fény esetén mozgás közben a Hue értéke változott. A sárga jelzésekép esetében jellemző volt, hogy a palettán mellette lévő piros értékbe ért bele.

A fentiek alapján úgy találtam, hogy egy-egy jól meghatározott színű LED-es fényforrás a Hue kis szórása mellett jó előzetes eredményt ad. Mivel a kísérleteket élő környezetben végeztem, valódi jelzőlámpákkal, a szórást befolyásolta a lencse, melynek alapszíne az alacsonyabb V (Value) érték mellett érvényesülni tudott. A modellemben egy olcsó kompakt kamera képes volt az 50 méterre lévő, $r = 2$ pixel sugarú jelzőlámpa fényének megtalálására, viszonylag nehéz textúrázott körülmények között is (47. ábra).

4.4.1 Összehasonlítás ma elterjedt megoldásokkal

A jelzőlámpa felismerésnél a state-of-the-art rendszerekre kutatásomat követően is a statikus, képkockán való felismerés volt a jellemző [111]. Számos módszer heurisztikus, vagy tanulási alapú modelleket mutatott be, de vannak olyanok is, melyek a teljesítmény javításához felhasználják a GPS koordinátát, térképeket is. A képfeldolgozásra épülő algoritmusok között látszólag nincs egyetértés abban, hogy melyik szintért tartja a felismerés szempontjából a legjobbnak, ebben szerepet játszhat az is, hogy a ráépülő megoldás milyen tulajdonságok mentén működik.

Hasonlóan az én megfigyelésemhez, a nagysebességű kamera alapú felismerését felhasználó [111] is 100Hz-es frekvencia melletti analízissel azonosít jelzésiképet, további tanító algoritmusok segítségével, egy erősnek tekinthető számítógépen C++-os kóddal kb. 7 fps sebességgel képes a frekvenciát a *Temporal Ternary Pattern* eljárás segítségével. A saját megoldásom MATLAB-os környezetben 3-4fps sebességgel volt erre képes, ahol a piros lámpa detektálása esetén ezzel az egyszerű összeállítással:

Összehasonlítva ezt más publikált és kidolgozottabb eljárásokkal ugyan rosszabb, viszont a mérési környezetben használt olcsó kamera miatt az artifakt hatás és az expozíciós idő nem állíthatósága miatt nehéz, kalibrálatlan környezetben készültek a felvételek.

TABLE II
AVERAGE PRECISION, RECALL, AND F_1 SCORE COMPARISON.

	<i>Precision</i>	<i>Recall</i>	F_1 <i>Score</i>
<i>proposed</i>	0.8419	0.9255	0.8719
MUL	0.4246	0.7996	0.4784
HAAR	0.4430	0.5304	0.4364
LBP	0.4243	0.4539	0.3913

48. ábra Nagy képsebesség melletti frekvencia felismerés alapján működő algoritmus eredménye (első sor) [111]

TABLE 6 Performance test for DL method.

Method	Recall	Precision	Frequency
Our method	0.86	0.87	27.8 FPS
Faster-RCNN	0.04	0.92	1.2 FPS
SSD	0.01	0.83	3.4 FPS
R-FCN	0.03	0.89	1.6 FPS

49. ábra Egy valós idejű, ConvNetet használó detektáló algoritmus eredménye (első sor) [104]

Az általam publikált megoldás a hasonló [111]-től egyértelműen elmarad, melynek oka, hogy a szegmentálás része nem lett elegendően jól kidolgozva, sokkal inkább a módszer alkalmasságán volt a hangsúly. Az ott publikált eredmény ráadásul egyéb osztályozási eljárást is magában foglal, viszont ez alapján látható, hogy nem feltétlenül szükséges nagysebességű képfelvétel készítés a frekvenciaelemzéshez. A valós idejű ConvNetet használó megoldás előhívási algoritmus viszont rosszabb, mint a frekvenciaanalízisen alapulón, vagyis a megtalálás esetén az előhívás frekvencia alapon hatékonyabb lehet.

4.4.2 A kutatás tárgyát képező modell felhasználásának lehetőségei

A modell egyszerű, álló helyzetben az elvártak szerint működött és a Goertzel algoritmus segítségével valóban felismerhetővé vált az adott térrészen lévő impulzus üzemű lámpa, így ez a megközelítés egy ilyen jól definiált problémaosztályra, miszerint alul-mintavételezetten, ismert frekvenciájú képi változás felismerése lehetséges. Az irodalomban az újabb eredmények alátámasztották, hogy a módszerrel alapvetően jelenleg akár jobb eredmény érhető el a ConvNetes megoldásokhoz képest a LED-es felismerésre, így érdemes lehet azt megnézni, hogy a frekvenciák a mély tanuló algoritmusok segítségével hogyan segíthetnek a képfeldolgozás felismerési feladatainak megoldásában.

A jelenség vizsgálatához kapcsolódó publikációm: [117]

5 Tudományos eredmények összefoglalása

5.1 Kutatás módszere

Kutatásom Roska Botond svájci kutatócsoportjának munkáját és eredményeit megismerve indult el 2008 végén, melyben a Celluláris Neurális Hálózatok (CNN) egyik komoly alkalmazási lehetőségét láttam meg. Először a kutatócsoport által megtalált retinacsatorna egyszerű modellezésével és a modell vizsgálatával foglalkoztam. A modellezési alapkérdéseket 2009-es diplomatervemben kezdtem el feldolgozni, melyben felvázoltam már a lehetséges kutatási irányokat, előzetes méréseket. 2009-ben, a Nature-ben megjelent cikkük adta meg a modell építésének konkrét sarokköveit és a modell behatóbb vizsgálatának lehetőségét, valamint ez alapján indulhatott meg a modell validációja és kalibrációja.

A modellt elemző munka alapvető fontosságú volt, mely során először a retinacsatorna működésének mélyebb megismerése, megértése volt az elsődleges feladat, valamint megfogalmaztam az egyes paraméterek meghatározásának kérdéseit. Ez alapján lehetett megkezdeni a modell validálását, valamint ehhez kapcsolódóan a modell kalibrálását, mely során vizsgáltam a modell működési tartományát és kapcsolódó időzítési viszonyait.

A modell felépítése során rendelkezésre álló kutatási eszközök lehetőséget adtak arra, hogy a modell alkotáshoz CNN chipen történő implementáció készüljön, így megvizsgáltam az algoritmus ezen architektúrán való megvalósíthatóságát. Az algoritmust a spanyol Anafocus Eye-RIS rendszerén készítettem el.

A modell vizsgálatának következő kérdésköre a kalibráció tisztázása volt, a modell matematikai alátámasztásával. A modell kalibrálásával kapcsolatos megállapítások 2012-ben jelentek meg referált folyóirat publikáció keretében. Ezen publikáció már az egér retina alapú modell más biológiai alapú modellekkel való összevetését is tartalmazta.

A modell vizsgálata során több, kisebb jelentőségű hatást is felismertem, például a közlekedési irány meghatározhatóságának kérdéskörét, valamint a kalibrációs paraméterek megváltoztatásai során észlelt másodlagos hatásokat és ezeket elemeztem is. Ez vezetett oda, hogy végül az impulzusüzemű LED-es fényforrások felismerésének lehetőségével foglalkoztam, ezen belül is a közlekedési jelzőlámpa felismerésén alul-mintavételezett frekvenciaanalízis segítségével.

5.2 Új tudományos eredmények

Munkám új tudományos eredményeit két fő téziscsoportba soroltam. Az első fő téziscsoportba azon téziseim kerültek, melyek közvetlenül az egérben talált Pvalb-5 ganglionsejt csoporthoz kapcsolódó modell kutatási eredményei alapján születtek meg. A második téziscsoportba a modellvizsgálat során talált egy mellékhatás kapcsán született eredmények kerültek. A disszertációhoz kapcsolódó tudományos eredmények folyóiratban [J], könyvfejezetben [B] és referált konferencia publikációkban [C] jelentek meg.

I. Téziscsoport:

1. Igazoltam, hogy az egerek retinájában felfedezett Pvalb-5 ganglionsejtek modellje képes az OFF receptív mezőn sötét tárgyak közeledését detektálni. Elemeztem a modell hatékonyságát implementálási szempontból. A modell viselkedési analízise kapcsán megmutattam, hogy az objektumok sebessége és iránya becsülhető.

Téziscsoporthoz kapcsolódó publikációm: [B], [J], [C1], [C2], [C3]

Az egér retinájában felfedezett Pvalb-5 ganglionsejt olyan OFF típusú sejt, mely érzékeny az objektumok közeledő mozgására. Egy Pvalb-5 ganglionsejt a látómező mintegy 10%-ának receptívmezőjét alkotó csap receptorsejtekből nyeri bemeneti jeleit. Egy receptív mező esetében a világos háttér előtt a retina felé közeledő sötétebb objektum sziluett növekedése nyomán előállt változás képes aktiválni a ganglionsejtet. A kísérletek bizonyították azt is, hogy a Pvalb-5 sejt ON irányban erősen gátló választ ad. Ennek jelentősége abban áll, hogy ezzel a sejt a laterális mozgásokra így teljesen érzéketlen, vagyis neuron választ nem ad. Az elkészített modellem ezt a működést bizonyította.

A csapok a modell szempontjából egyszerre viselkednek serkentő és gátló sejtneként, függően attól, hogy OFF vagy ON típusú eseményt érzékelnek. Az egyes receptorsejteken beérkező stimulus, függően attól, hogy serkentőként, vagy gátlóként viselkedik az előtte történetekre, a Pvalb-5 ganglionsejten összegződik, mely neurális sejt választ ad.

Az elkészült modell információ áramlását a 23. ábra mutatja be, melyen az látható, hogy valamely receptorsejten beérkező $s(t)$ időbeli stimulusra (6) szerint számolható a serkentő és (7) szerint a gátló időbeli konvolúció. A serkentés és a gátlás esetén egyaránt lineáris szűrést kell végezni a (8) és (9) szerint számolható módon. Ennek jelentősége abban van, hogy a természetben csak “pozitív” sejt válasz létezik, ez a lépés ezt készíti elő. Emiatt van szerepe a gátló csatorna hatásnak.

A kialakult lokális eredmények a Pvalb-5 sejt által lefedett receptívmező összegzi, s ennek küszöbérték fölötti része adja meg a becsült sejtválaszt a közeledés tényére, melyet a (10) ír le.

A modellt megvizsgálva látható volt, hogy a közeledést vizsgáló esetekben a receptívmező változásainál pozitív előjellel vett változás történik (serkentés), ahol nincs történés, ott semleges. A laterális mozgásnál viszont a serkentő változással egyidőben gátló változás is megjelenik.

Tézis 1.1: Megmutattam, hogy az egerek retinájában felfedezett, Pvalb-5 típusú ganglionsejtnél leírt működés párhuzamos architektúrákon megvalósítható és hatékonyan implementálható.

Tézishez kapcsolódó publikáció: [B], [J], [C1]

A modellt leíró (6)-(10) képleteknél az (6) -(9) képletek egy pixelre elvégzett számítások, melyek egyszerű szorzások és összeadások, valamint egy szignum függvény alkalmazása. A (10) képlet a térbeli összegzést valósítja meg. Mint a képletekből látható, elemi műveletekkel egyszerűen leírható a modell és a pixel szintű műveletek hatékonyan párhuzamosíthatók SIMD architektúrára. A térbeli összegzés a pixelek számának és a rendelkezésre álló processzormagok számának, valamint a processzor architektúrális felépítésének függvénye.

Tézis 1.2: Kísérletileg bizonyítottam, hogy az egerek retinájában felfedezett, Pvalb-5 típusú ganglionsejtnél leírt működés topografikus processzor architektúrára elemi CNN műveletekkel megvalósítható, és analóg processzor tömbön hatékonyan implementálható sebesség-energia-terület viszonylatban.

Tézishez kapcsolódó publikáció: [B], [J], [C1]

A Pvalb-5 ganglionsejt a látótérből 10° -ot fed le, ami nagyszámú csap stimulusának feldolgozását jelenti. Az egér látómezeje $120-130^\circ$ körüli, ezt százezres nagyságrendű csap fedi le, a különböző típusú csapok eloszlása a látómezőn nem egyenletes, kb. 15% egységnyi felületen a sűrűségben az eltérés.

A modellezéshez az Eye-RIS v1.2 CNN-UM analóg érzékelő számítógépre esett a választásom, mely a kutatás idején az egyik legkorszerűbb fókuszsíkbeli analóg processzortömbbel rendelkező érzékelő-számítógép. A rendszer egy Q-Eye szenzor-processzor chipet tartalmaz, mely szürkeárnyalatos felvételek készítésére akár 10000 fps sebességgel képes 176x144-es képméretben. A CNN műveletek végrehajtását alapértelmezés szerint saját FPP nevű funkcionális programozási nyelvvel utasításszinten támogatja. Egy parancs lefuttatása során kizárólag az analóg végrehajtási idővel kell számolni.

Az implementáció a ganglionsejt kimenetét az Eye-RIS-on 58-98 μ s alatt képes meghatározni a képkészítést követően. A 30. ábra bemutatja az egyes műveletek részidejeit.

A ganglionsejt lényegében egy morfológiai összegzést végez, ez a legdrágább művelet az egész folyamat során. Ez az összegzés lényegében egy diffúzió, melyet az Eye-RIS a képtől és egyéb beállításoktól függően 10-50 μ s alatt tud végrehajtani. A digitális architektúrákhoz képest itt az időben számottevő nyereséget lehet elérni. Az analóg diffúzió ideje függ a háttértől, zajoktól és a beállításoktól. Ez a diffúz összegzés valósítja meg a (11)-es képlet szerint leírtakat.

Tézis 1.3: Megmutattam, hogy több Pvalb-5 típusú ganglionsejt a retina különböző térrészein különböző módon reagál ugyanazon gyorsan közeledő objektumra, legerősebben a közeledés centrális közepén reagál.

Tézishez kapcsolódó publikációm: [B], [C3]

Roska Botond kutatócsoportja bemutatta [6], hogy a ganglionsejtek által lefedett területek átfedésben vannak egymással, átlapolással fedik le a receptívmezőt. A modell módosításával átfedő hálózatot hoztam létre és megmutattam, hogy a receptívmező által látott méretét meg nem haladó látótérnövekedésben a középső receptívmezőre hat a leginkább és az átfedő részekre kevésbé, viszont a receptívmezőt kinőve a szomszédos mezők még továbbra is reagáltak, így nőtt az aktivitás mértéke.

Tézis 1.4. Karakterizáltam és kalibráltam a Pvalb-5 típusú ganglionsejt működését leíró modellt sebesség-méret-távolság relációkban

Tézishez kapcsolódó publikáció: [B], [J]

A karakterizálás négy fő paraméter megadását jelenti:

1. Konvolúciós ablakok nagysága (a rendszer memóriája, reagálási képessége a közeledés sebességére)
2. A serkentő és gátló csatornák küszöbértékei (kis változások és zajok eliminálásának képessége, a méretváltozásra való érzékenység)
3. Receptív mező mérete (a közeledő objektum méretéhez, távolságához képest, a lehetséges sejtválasz nagyságát befolyásolja)
4. Ganglionsejt küszöbértékének meghatározása (a modell jelzési küszöbe, az előző paraméterek függvényében)

Az átfedő receptívmezők és méretezésük kapcsán megvizsgáltam, hogy a receptívmező mérete, a közeledő objektum mérete és sebessége, valamint távolsága, irányultsága milyen összefüggésben befolyásolja a Pvalb-5 ganglionsejt választát. Közeledő, távolodó, valamint laterális mozgások és ezek kombinációival modellkísérleteket végeztem, valamint felírtam a modellre jellemző alapösszefüggéseket, valamint elemeztem a modell paraméterezését ezek függvényében.

A modell számítást plotterasztal segítségével validáltam, különböző sebességű és irányultságú mozgások vizsgálatán keresztül.

Tézis 1.5: Megmutattam, hogy a Pvalb-5 típusú ganglionsejt működése alapján, egy közeledő objektum iránya becsülhető, valamint a világos háttéren sötét objektumok laterális elmozdulásának irányát is képes megmutatni.

Tézishez kapcsolódó publikáció: [B], [C3]

Mivel a közeledő objektum az átfedő ganglionsejteket különböző mértékben aktíválja, így ezen aktivitási mintázat a ganglionsejt válaszána szignumfüggvényével vett küszöbölését követően, meghatározható a közeledés irányának vektora, az aktivitás mértékének függvényében, az átfedő receptormezők pozíciójához képest.

Ezen vektor lényegét tekintve egy receptívmező és annak szomszédos átfedő receptívmezőinek szuperpozíciójából adódik, mely a (19)-el számolható, mely a konkrét 41. ábra szerinti átfedő receptívmezők esetén (20) megoldását jelenti. A modellről megmutattam, hogy képes jelezni a laterális mozgásokat is.

II. Téziscsoport:

2. Megmutattam, hogy a periódikusan változó fényintenzitást kibocsájtó impulzus üzemű LED-es lámpák időbeli alul-mintavételezéssel detektálhatóak HSV színtartománnyal történő szegmentációt követően, mozgó- illetve álló platformos képfelvétel esetén.

Tézishez kapcsolódó publikációm: [C4]

A közeledésre érzékeny retinacsatorna vizsgálata során kiderült, hogy a periodikusan változó intenzitású LED-es fényforrások alacsony képrögzítési sebesség mellett a receptívmező modell számolása során a kimenet végeredményét befolyásolják. Ennek oka, hogy folyamatos ON-OFF impulzus éri egy adott területen a receptívmezőt és ezzel a pixel szintű kezdeti konvolúciót befolyásolja. Megfelelő színszegmentációt követően a frekvencia alapján egy jelzőlámpa felismerhető.

A szegmentáció során fontos szempont, hogy minél szűkebb, de az összes lehetséges találatot tartalmazó tartomány ki legyen jelölve. A LED-ek látható fényhullámhossza szűk tartományban van, így szegmentációjuk egyszerűbb, jelzőlámpák esetén pedig szabványok is vonatkoznak a lámpák színére és a fény erősségére. A LED-es lámpáknál az emittált fény színe konstans és csak a fényereje, valamint szaturációja változik.

Megvizsgáltam a különböző szintereket a jelzőlámpa felismerés problémaköre kapcsán, s úgy találtam, hogy a legalkalmasabban a HSV színtér alkalmazható szegmentációs feladatra. A H (Hue) értéke jól meghatározottnak tekinthető a szabványok miatt, míg a másik két paraméter függ a környezettől és a képképző eljárástól.

A LED egyenletes ütemű, impulzus szerű villogása alacsony képsebesség mellett a jelenség alul-mintavételezését jelenti, melynek frekvenciameghatározása a Nyquist-Shannon mintavételezési tétel miatt nem lehetséges. Kihasználva viszont, hogy állandó a villogás

frekvenciája, így a Goertzel algoritmussal, megfelelő mintaszám esetén még alulminavételezés mellett is következtethetünk impulzusüzemű fényforrás, így jelzőlámpa felismerésére.

Az egyszerű színszegmentációs algoritmust követően látható volt, hogy az algoritmus képes volt a frekvencia alapján a felismerési feladatra.

Tézis 2.1: Adható olyan analóg CNN processzortömbön is implementálható megoldás a frekvencia számolására, mely képes impulzus üzemű fényforrások detektálására.

A Goertzel algoritmus előnyös tulajdonsága, hogy egyszerűen számolható pixel szinten, s elegendő mennyiségű analóg memória mellett:

```

realW = 2*cos(2*pi*k/N); //FPGA művelet cos miatt
imagW = Sin(2*pi*k/N); //FPGA művelet sin miatt

LAM_1 = 0; //CNN művelet
LAM_2 = 0; //CNN művelet
LAM_0 = 0; //CNN művelet
for (n=0; n<N; ++n) //ciklus FPGA közreműködéssel
{
    LAM_0 = LAM_x(n)+(realW*LAM_1)-LAM_2; //CNN művelet
    LAM_2 = LAM_1; //CNN művelet
    LAM_1 = LAM_0; //CNN művelet
}
outreal = 0.5*realW * LAM_1 - LAM_2; //CNN művelet
outimag = imagW * LAM_1; //CNN művelet

```

A Goertzel algoritmust megvalósító kódban a ciklusban található műveletek a LAM analóg memóriacellák között értelmezett műveletek, a ciklus hívások pedig az FPGA-ról kezdeményezhetők, ahogy a szinusz és koszinusz számolás is. Az Eye-RIS v1.3 rendszeren a fenti algoritmus valósídejű alkalmazásként nem implementálható, mivel az analóg processzortömb 6 analóg memóriacellája nem elegendő a számolások elvégzéséhez.

5.3 A képfeldolgozási lánc és biológiai vonatkozásai, a kutatás eredményeinek felhasználása, aktualitása

A közlekedés-informatikában legyen az vasút, autóipar, vagy éppen repülés, nagy mennyiségű adat feldolgozására van igény. Mindhárom rendszerre igaz, hogy teljesen autonóm rendszer még jelenleg nem létezik, talán legelőrébb a vasút áll, mely zárt és jól karakterizált pálya esetén (CBTC alapú metrók) képes alacsony sebesség mellett akár tízméteres nagyságrendű követési távolságot betartva is egymás után szerelvényeket közlekedtetni radar, vagy más a távolságot aktívan mérő technológia nélkül, pusztán helyadatok és sebességmérés, valamint a saját állapotának ismeretében (sebesség, megfékezettesség). A “nagyvasúton”, mely nem zárt rendszer, mint a metró, valós probléma az objektumok pályára kerülésének figyelése, akkor is, ha az ütközés nem elkerülhető, mert ilyenkor egy nagyobb tárgyal való ütközés előtti figyelmeztetésre a mozdonyvezetők a mozdonytér védettebb területére tudnak menekülni.

A repülőök autonóm vezetéséhez nagymennyiségű szenzort vetnek be, melyek egy része az objektumérzékelések kapcsán aktív (radarok), de az autonóm repülés esetében (UAV) számtalan kutatásban kerül elő a kamerák felhasználásának kérdése is az objektumészlelésnél.

A vezetést támogató rendszerek kapcsán a ma viszonylag drága és kevésbé praktikus, aktív eszközök, mint a radar és LIDAR megoldások mellett az elmúlt 15 évben számtalan cég foglalkozott kamerás rendszerekre épülő megoldásokkal. Az autonóm vezetésben a közeledő objektumok felismerése kritikus és állandóan jelenlévő probléma, melyre hatékony megoldást kell adni. Jelenleg a legtöbb megoldás a közeledő objektumok érzékelését egyedül kamerára még nem meri bízni, így a most piacon lévő egyik legfejlettebbnek tekinthető beágyazott EyeQ4 chip megoldás is tartalmazza a radarral együtt való alkalmazás lehetőségét [51], pedig maga a chip alkalmas biztonságkritikus számítási feladatok ellátására.

A figyelmeztető rendszerek esetében fontos a hibás pozitív jelzések minimálisra szorítása, míg az autonóm rendszereknél mindemellett szükséges a nagy biztonságú felismerés is, mely csak meghatározottan tolerálható szintű hibázást tesz lehetővé.

A gyors és pontos felismeréshez mindenképpen fontos az, objektum közeledés tényét időben észlelni, hogy a figyelmeztetési vagy beavatkozási döntésre több idő juthasson.

A kutatásom óta eltelt időben egyértelmű tendencia volt a klasszikus, feladatorientált szegmentáció háttérbeszorulása és a mély tanuló algoritmusok, valamint hatékony feldolgozási hálózatok előretörése.

Ezekben az élő szervezetekhez hasonló megközelítésűek már, de mint látható, azzal a kivétellel, hogy bizonyos fontos képfeldolgozási műveletek nem magas (agyi) szinten kerülnek felismerésre, hanem már alacsony szinten (retina), mint a közeledő objektum esetében.

A sáskában felfedezett, a Pvalb-5-nél jóval közismertebb és durvább felbontású LGMD1 modellel mind a mai napig többen foglalkoznak a robotikában utolsó pillanatos ütközés-elkerülési megoldások kapcsán, ennek vélhető oka, hogy chip implementációja is létezik. Figyelemre méltó, hogy a precízebb felismerést támogató LGMD2 megoldás sem terjedt el a Pvalb-5-höz hasonlóan.

A Pvalb-5-öt számára tipikusan a jövőbeli alkalmazási terület lehet az érzékelő szenzor-processzortömbök további fejlődése és szélesebb körben való elterjedése, melyek a szintén biológiai környezetet inspirálta ConvNetek számára is alkalmas futási környezetet jelenthetnek.

A lámpa jelzésekép felismerése kapcsán a frekvencia elemzés alapján történő felismerést nagysebességű felvételeken mások is kutatták a publikációmát követően. Az erre a felismerésre építő algoritmusok a felismerésben az elmúlt évben született ConvNetes lámpa felismerési eljárásokhoz képest jobb eredményt mutatnak. Abban továbbra is egyedinek tekinthető a publikált eljárásom, hogy alul-mintavételezéssel való megközelítéssel, hogy hagyományosabb képfelvevő eszközökkel és számítási apparátussal oldható meg vele a fényforrások felismerési feladata, ez pedig a számítási feladatokban is előnyösebbé teszi. A Goertzel algoritmus mindemellett alkalmas analóg CNN eszközön implementálható.

Eredményeim rámutatnak arra is, hogy a természetben számos problémára van megoldás, így utat kíván mutatni annak is, hogy a természettől mind több el lehet lesni és megvalósítani.

6 Publikációs lista

Könyvfejezet

[B] Á. Zarándy és T. Fülöp, „Implementation and Validation of a Looming Object Detector Model Derived from Mammalian Retinal Circuit,” in *Focal-Plane Sensor-Processor Chips*, Springer, 2011, pp. 245-260.

Folyóirat

[J] Á. Zarándy és T. Fülöp, „Approaching object detector mouse retina circuit model analysis and implementation on cellular sensor-processor array,” *INTERNATIONAL JOURNAL OF CIRCUIT THEORY AND APPLICATIONS*, Vol. 40., pp. 1249-1264., 2012.

Konferencia

[C1] T. Fülöp és Á. Zarándy, „Bio-inspired looming object detector algorithm on the Eye-RIS focal plane-processor system,” in *12th international workshop on cellular nanoscale networks and their applications - CNNA*, Berkeley, 2010.

[C2] Á. Zarándy és T. Fülöp, „Retinal approaching object detector model implementation and validation,” in *IEEE international symposium on circuits and systems - ISCAS*, Paris, 2010.

[C3] T. Fülöp és Á. Zarándy, „Bio-inspired looming direction detection method,” in *13th International Workshop on Cellular Nanoscale Networks and their Applications - CNNA*, Turin, 2012.

[C4] T. Fülöp és Á. Zarándy, „Chromatic Spatial-temporal Frequency Analysis for Traffic Light detection,” in *4th IEEE International Conference on Cognitive Infocommunications - CogInfoCom*, Budapest, 2013.

[C5] T. Fülöp és Á. Zarándy, „Real-time moving object segmentation algorithm implemented on the Eye-RIS focal plane sensor-processor system,” in *International symposium on nonlinear theory and its applications - NOLTA*, Budapest, 2008.

7 Köszönet

Sok türelmet és támogatást köszönhetek sokaknak. Azért, hogy eljutottam, eljuthattam idáig. Öröm, hogy ennyi jó ember vesz körül.

Ákosnak köszönöm a sok iránymutatást, javaslatot, szakmai és személyes beszélgetést, biztatást, munkáim átnézését, vagy éppen angol írásaim végigszenvedését, amivel 2007 óta feltétlenül támogatott. Közös munkánk kapcsán nem tudom nem hivatkozni *Lukács 15,11–32-t*. Köszönöm mindenkori és feltétlen támogatásodat.

Ákos mellett Vető István Tanár úr az, akinek nagyon sokat köszönhetek mind szakmai, mind emberi oldalról egyaránt, a sok egyetem környékén töltött év alatt. Értékes beszélgetéseink nagyon hiányoznak. Embersége, szakmaisága, mérnökséghez való hozzáállása és mindenkor bátorító, jóindulatú és alázatos hozzáállása követendő és továbbadandó példaként áll számomra.

Roska Tamás professzor úrnak köszönhetem, hogy 2007-ben a SZTAKI-ban, majd 2009-ben már a doktori iskolában lehetőséget adott, hogy megismerjem a kutatói életutat, megtudjam, hogy mit jelent az elmélyült kutatás az alázatos munka, lelkesen hozzáállni egy-egy feladathoz. A disszertáció megírásakor ezt újra átélhettem, s ez most vissza hozta a kedvet, hogy ne szakadjak el teljesen a továbbiakban sem az egyetemi és kutatói körtől, s lássak benne perspektívát a valamikori jövőben.

Köszönöm Szolgay Péter és Vida Tivadarné folyamatos támogatását, mindenkori jelenlétét és rám szánt figyelmét, intő szavaikat. Ez tette lehetővé, hogy megembereljem magam, és végül a befejezés mellett döntsek.

Köszönöm, hogy Vágó Zsuzsanna megbuktatott Analízis II-ből a harmadik vizsgán is. “Én úgy látom, hogy maga ezt nem érti.” szavai máig fülemben csengenek. Akkor és ott helyén való volt: emlékeztetett újra Apum és általános iskolai kémia tanárom, Kaposiné Héger Ágnes szavaira, akik mindig arra tanítottak, hogy értsük meg a dolgokat, a világot, az összefüggéseket.

A Kalmárból Pataky Margit tanárnőnek köszönöm azokat a stabil és mély számítástechnikai szakmai alapokat, melyekkel megalapozta szakmai hozzáállásom.

Szombatfalvy Csillának, hogy megtanított írni, a komolyzene megszerettetésével együtt: Az orosz ötöktől, Esterházy Pál Harmonia caelestisén át Krzysztof Pendereckiig.

Köszönöm a sok év közös oktatói munkát, bizalmat és biztatást Bércesné Novák Ágnes tanárnőnek, valamint Lukács Gergelynek, akinél TDK-mal a kutatói utat először kipróbálhattam. Oláh Andrásnak is köszönetet mondok, akivel az egyetemi szakmai versenyélet és utánpótlás gondozás elindításában dolgozhattam együtt. Köszönettel tartozom Tihanyi Attilának és Péri Mártonnak, akik álltak rendelkezésemre bármikor, amikor szakterületükről segítséget kértem.

Hálás vagyok mindenkori egyetemi kollégáimnak. Különösen, akik részt vettek és mindig társaim voltak doktoranduszként a mindennapokban, gyakorlatok tartásánál, akikkel beszélhettem szakmai és társadalmi kérdésekről, nehézségekről, vagy épp eredményekről, meglátásaimról: Radványi Mihály, Koller Miklós, Tornai Kálmán, Füredi László, Stubendek Attila, Tornai Gábor, Horváth András, Kiss András, Tisza Dávid, Tóth Emília, Kozák László. Kiemelném Zsedrovits Tamást, akinek számtalan közös gyakorlatvezetést, beszélgetést és angol lektorálást köszönhetek az elmúlt bő tíz évben.

Köszönöm, hogy belekóstolhattam az ipari kutatásba is a FETI-nél, a közös munkát, különösen Kováts Antalnak, Soós Balázs Gergőnek és Aoyagi-sannak.

Köszönöm jelenlegi munkahelyem, a Prolan Irányítástechnikai Zrt. támogatását, kollégáim türelmét és érdeklődését, hogy időt adtak arra, hogy mindennapi feladataim mellett nyugodtan tudjam megírni disszertációm is.

Külön köszönöm mindazok támogatását, akik végig mellettem voltak, biztattak, érdeklődtek. Legelőször és legjobban Orsinak, akivel életem akkor fonódott össze, amikor neki kezdtem ennek a kalandnak és akkor lett igazán egy, mikor épp a befejezéshez készülök már. Nem kevés türelmét köszönöm.

Szüleimnek. Mindhárom testvéremnek. Keresztszüleimnek.

Legjobb barátaimnak is köszönöm türelmüket és támogatásukat: Balázsnak, Csongornak, Ferinek, Gergőnek. Közeli barátaimnak is a folyamatos érdeklődést, biztatást: Bálintnak, Dávidnak, Eriknek, Fenyőfalvi Tamásnak, KV-nak, Bálint Zolinak, Mandácskó Zolinak és sógoromnak, Ciánnak.

A MAVÉ-s és EESTEC-es barátaimnak, külön kiemelve Franciskát, Ticz Ádámot, Krisztát, Krisztiánt, Renit, Csabit, akik mindannyian többször/sokszor kérdezték találkozáink alkalmával, hogy hogy állok velem.

A PhD disszertációt az Európai Unió támogatásával valósítottam meg, felhasználva az Európai Szociális Alap támogatását (EFOP -3.6.3-VEKOP-16-2017-00002)

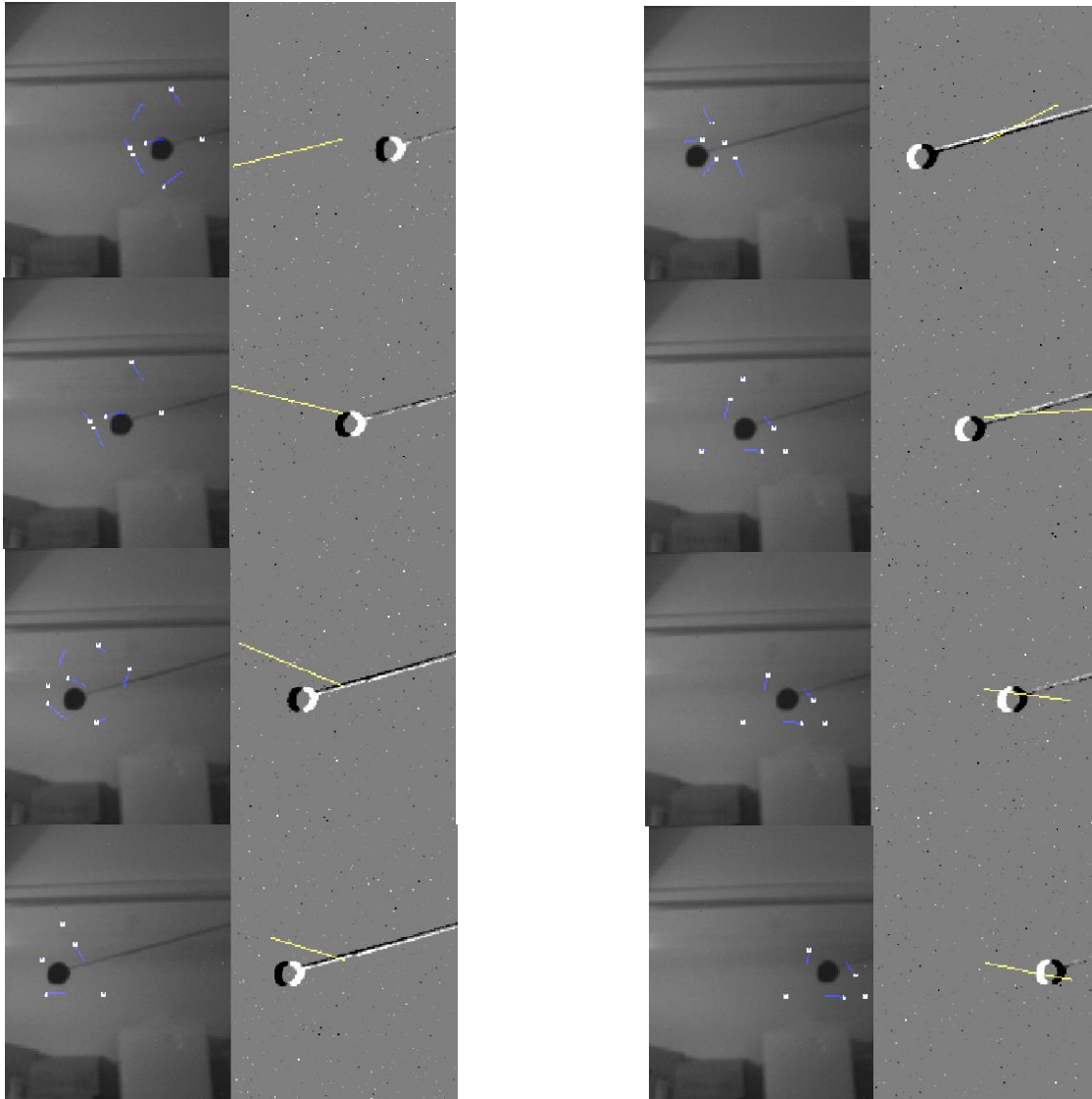
8 Rövidítések jegyzéke

ADAS	Advanced Driver-Assistance Systems	Vezetőt támogató rendszerek gyűjtőneve
cGMP	cyclic Guanosine Monophosphate	Ciklikus guanozin-monofoszfát: kémiai hírvivő molekula
ConvNet	Convolutional Neural Network	Konvolúciós neurális hálózat
CNC	Computer Numerical Control	Precíz, számítógép vezérlésű robotkaros szerszámgépek elnevezése.
CNN	Cellular Neural Network	Celluláris neurális hálózat
CNN-UM	CNN Universal Machine	CNN univerzális gép
FOE	Focus of Expansion	Növekedési fókuszpont: egy olyan látszólagos pont, ahonnan egy objektum közeledés indul
GPGPU	General Purpose Graphic Processing Unit	Általános célú grafikus processzor
HDR	High Dynamic Range Imaging	Nagy dinamikatarományú képalkotási eljárás
HOG	Histogram of Oriented Gradients	Gradiens irány hisztogram: kép tulajdonság leíró eljárás
LED	Light-Emitting Diode	Fényt kibocsátó dióda
LG	Nucleus Geniculatus Lateralis	oldalsó geniculatus mag: agyban található sejtcsoport
LGMD	Lobula Giant Movement Detector	Lobula óriás mozgásérzékelő látó-neuron
LIDAR	Light Detection and Ranging	Lézer alapú távérzékelő
RADAR	Radio Detection And Ranging	Rádió alapú távérzékelő
ROI	Region of Interest	Érdekes régió: valamilyen szegmentáció során kijelölt, a további

		feldolgozás szempontjából fontos képterület
SIFT	Scale Invariant Feature Transform	Invariáns jellemző pont detektor: kép tulajdonság leíró eljárás
SIMD	Single Instruction Multiple Data	Egy utasítás több adaton: architektúrális elrendezés, mely során egy kiadott utasításra a műveletet párhuzamosan több adaton végzi el a processzoregység.
SVM	Support Vector Machine	Szupport Vektor Gép/Tartóvektorgép: gépi tanulásnál egy felügyelt tanulási modell osztályozási feladatokra

9 Függelék

9.1 Balra és jobbra történő laterális mozgásnál jelzett irány, valamint mozgó platformon lokális elmozdulás vektorok



10 Hivatkozások

- [1] P. I. Pavlov, „Conditioned Reflexes. An Investigation of the Physiological Activity of the Cerebral Cortex,” *The Journal of Nervous and Mental Disease*, Vol. 69., pp. 470-471, 1929.
- [2] J. Szentágothai és M. Réthelyi, „Gerincevelői reflexívek,” in *Funkcionális anatómia*, Budapest, Medicina Könyvkiadó, 2006., pp. 1403-1417.
- [3] R. Sekuler és R. Blake, „Észlelés,” Budapest, Osiris, 2000, pp. 21-44..
- [4] T. Fülöp és Á. Zarándy, „Real-time moving object segmentation algorithm implemented on the Eye-RIS focal plane sensor-processor system,” in *International symposium on nonlinear theory and its applications - NOLTA*, Budapest, 2008.
- [5] L. Alba, R. D. Castro, F. Jimenez-Garrido, S. Espejo, S. Morillas, J. Listan, C. Utrera, A. Garcia, M. D. Pardo, R. Romay, C. Mendoza, A. Jimenez és A. Rodriguez-Vazquez, „New Visual Sensors and Processors,” in *Spatial Temporal Patterns for Action-Oriented Perception in Roving Robots*, Heidelberg, Springer, 2009, pp. 351-369.
- [6] T. Münch, R. Silveira, S. Siegert, T. Viney, G. Awatramani és B. Roska, „Approach sensitivity in the retina processed by a multifunctional neural circuit,” *Nature Neuroscience*, Vol 12., pp. 1308-1316, 2009.
- [7] V. Csépe, M. Gyóri és A. Ragó, Általános pszichológia 1-3., Budapest: Osiris Kiadó, 2011..
- [8] J. J. Gibson, „Optical motions and transformations as stimuli for visual perception.,” *Psychological Review* Vol 64., pp. 288-296., 1957.
- [9] Wikimedia, „Wikipedia,” Wikimedia, 3 10 2018. [Online]. Available: https://hu.wikipedia.org/wiki/David_Hunter_Hubel. [Hozzáférés dátuma: 07 09 2019].
- [10] G. Riddoch, „On the Relative Perceptions of Movement and a Stationary Object in Certain Visual Disturbances due to Occipital Injuries,” *Proceedings of the Royal Society of Medicine*, pp. 13-34, 1917.

- [11] L. Weiskrantz, *Blindsight: a case study and implications*, Oxford: Oxford University Press, 1990.
- [12] L. Kardos, „A pszichológiában alkalmazott kutatásimódszerek elvi kérdései,” in *A pszichológia módszerei - Pszichológiai tanulmányok XII.*, Budapest, Akadémiai Kiadó, 1970, pp. 7-28.
- [13] G. Fehér, „A házimadarak funkcionális anatómiája - A látás szerve, organum visis,” in *A háziállatok funkcionális anatómiája*, Budapest, Mezőgazda kiadó, 2004., pp. 1323-1327..
- [14] J. Szentágothai és M. Réthelyi, „Látószerv,” in *Funcionális Anatómia*, Budapest, Medicina, 2006., pp. 1681-1731.
- [15] G. Fehér, „Szabályozókészülék - A látás szerve, organum visus,” in *A háziállatok funkcionális anatómiája*, Budapest, Mezőgazda Kiadó, 2004, pp. 1094-1110..
- [16] A. Weismann és J. Csemez, „A rovarok látása,” *Természettudományi közlöny, Vol. 27*, pp. 515-530, 1895.
- [17] J. L. Nation, *Insect Physiology and Biochemistry*, US: CRC Press, 2008, p. 321.
- [18] M. Wilson, P. Garrard és S. McGinness, „The Unit Structure of the Locust Compound Eye,” *Cell and Tissue Research, Vol 195.*, pp. 205-226, 1978.
- [19] Wikipedia, „Ommatidium,” Wikimedia, 07. 05. 2018.. [Online]. Available: <https://en.wikipedia.org/wiki/Ommatidium>. [Hozzáférés dátuma: 14. 05. 2018.].
- [20] J. Tóth, „A fény érzékelése,” in *Erdészeti rovartan*, Budapest, Agroinform és Educatio, 2014., pp. 36-38..
- [21] J. E. Dowling és B. B. Boycott, „Organization of the primate retina: electron microscopy,” *Proceedings of the Royal Society B: Biological Sciences, Vol. 166.*, pp. 80-111., 1966.
- [22] Wikimedia, „Wikipedia: Emberi szem,” Wikimedia, 30. 03. 2018.. [Online]. Available: https://hu.wikipedia.org/wiki/Emberi_szem. [Hozzáférés dátuma: 14. 05. 2018.].
- [23] R. Pál, *Retina*. [Performance]. Semmelweis Egyetem, ÁOK, Anatómiai, Szövet- és Fejlődéstani Intézet, 2017. 11. 08..

- [24] I. Kalló és R. Gábrriel, *Bevezetés a funkcionális neurobiológiába: Retina*. [Performance]. PPKE-ITK és SOTE, 2018. 02. 26..
- [25] I. Kalló, *Introduction to functional neurobiology: Visual processing*. [Performance]. PPKE-ITK, 2018.
- [26] S. Roy, „Reconstruction of a class of fluid flows by variational methods and inversion of integral transforms in tomography,” University of Texas, Arlington, 2015.
- [27] Wikimedia, „Gépi látás - Wikipedia,” Wikimedia, 05. 06. 2018. [Online]. Available: https://hu.wikipedia.org/wiki/Gépi_látás. [Hozzáférés dátuma: 16. 06. 2018].
- [28] P. N. Stuart Russel, „Az észlelés,” in *Mesterséges Intelligencia Modern megközelítésben*, Budapest, Panem, 2005, pp. 984-1020..
- [29] AIA, „Vision Online,” Global Association for Vision Information, 16. január 2014.. [Online]. Available: https://www.visiononline.org/vision-resources-details.cfm/vision-resources/Computer-Vision-vs-Machine-Vision/content_id/4585. [Hozzáférés dátuma: 16. március 2018].
- [30] B. Jähne, *Digital Image Processing - 6th revised and extended edition*, Heidelberg: Springer, 2005.
- [31] A. Rövid, Z. Vámosy és S. Sergyán, *A gépi látás és képfeldolgozás párhuzamos modelljei és algoritmusai*, Budapest: Typotex, 2014.
- [32] S. Saha, „A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way,” *Towards Data Science*, 15 12 2018. [Online]. Available: <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>. [Hozzáférés dátuma: 08 09 2019].
- [33] Z. Kató, *Optikaiáramlás és követés*. [Performance]. SZTE - Képfeldolgozás és Számítógépes Grafika tanszék, 2014.
- [34] L. O. Chua és L. Yang, „Cellular Neural Networks: Theory,” *IEEE Transactions on Circuits and Systems, Vol. 35.*, pp. 1257-1272, 1988.

- [35] J. J. Hopfield, „Neural networks and physical systems with emergent collective computational abilities,” *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 79., pp. 2554-2558., 1982.
- [36] Á. Zarándy, A. Horváth és P. Szolgay, „CNN Technology-Tools and Applications,” *IEEE Circuits and Systems Magazine*, Vol. 18, pp. 77-89, 2018.
- [37] D. Bálya, B. Roska, T. Roska és F. S. Werblin, „A CNN framework for modeling parallel processing in a mammalian retina,” *International Journal of Circuit Theory and Applications*, Vol. 30, pp. 363-393, 2002.
- [38] Z. Vörösházi, Z. Nagy és P. Szolgay, „FPGA-Based Real Time, Multichannel Emulated-DigitalRetina Model Implementation,” *EURASIP Journal on Advances in Signal Processing*, Vol. 6, 2009.
- [39] M. Stewart, „Simple Introduction to Convolutional Neural Networks,” 27 02 2019. [Online]. Available: <https://towardsdatascience.com/simple-introduction-to-convolutional-neural-networks-cdf8d3077bac>. [Hozzáférés dátuma: 08 09 2019].
- [40] A. Horváth, M. Hillmer, Q. Lou, X. S. Hu és M. Niemier, „Cellular Neural Network Friendly Convolutional Neural Networks – CNNs with CNNs,” in *Design, Automation & Test in Europe Conference & Exhibition*, Lausanne, 2017.
- [41] F. S. Werblin, T. Roska és L. O. Chua, „The analogic cellular neural network as a bionic eye,” *International Journal on Circuit Theory and Applications*, Vol 23., pp. 541-569, 1995.
- [42] F. Werblin, „Synaptic connections, receptive fields, and patterns of activity in the tiger salamander retina. A simulation of patterns of activity formed at each cellular level from photoreceptors to ganglion cells,” *Investigative Ophthalmology & Visual Science*, Vol. 32., pp. 459-483, 1991.
- [43] NVIDIA, „About CUDA | NVIDIA Developer,” NVIDIA Corporation, 2018. [Online]. Available: <https://developer.nvidia.com/about-cuda>. [Hozzáférés dátuma: 16. március 2018].
- [44] G. Linan, S. Espejo, R. Dominguez-Castro és A. Rodriguez-Vazquez, „ACE4k: an analog I/O 64x64 visual microprocessor chip with 7-bit analog accuracy,” *International Journal Circuit Theory and Applications*, Vol. 30., pp. 89-116, 2002.

- [45] A. Rodriguez-Vazquez, G. Linan-Cembrano, L. Carranza, E. Roca-Moreno, R. Carmona-Galan, F. Jimenez-Garrido, R. Dominguez-Castro és S. E. Meana, „ACE16k: the third generation of mixed-signal SIMD-CNN ACE chips toward VSoCs,” *IEEE Trans. Circuits and Systems*, Vol. 51., pp. 851-863, 2004.
- [46] P. Földesy, Á. Zarándy, C. Rekeczky és T. Roska, „Digital implementation of cellular sensor-computers,” *INTERNATIONAL JOURNAL OF CIRCUIT THEORY AND APPLICATIONS*, Vol. 34., pp. 409-428, 2006.
- [47] L. O. Chua és T. Roska, „The CNN Universal Machine,” 1992.
- [48] T. Roska és L. O. Chua, „The CNN universal machine: an analogic array computer,” *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, Vol. 40., pp. 163-173., 1993.
- [49] C. Pan és A. Naeemi, „A Proposal for Energy-Efficient Cellular Neural Network Based on Spintronic Devices,” *IEEE Transactions on Nanotechnology*, Vol. 15, pp. 820-827, 2017.
- [50] C. Pan és A. Naeemi, „Non-Boolean Computing Benchmarking for beyond-CMOS Devices based on Cellular Neural Network,” *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits*, Vol. 2, pp. 36-43, 2016.
- [51] e-Hike, „e-Hike,” e-Hike, 2018. [Online]. Available: <http://www.e-hike.net/tr/content/mobileye-unveils-gen4-system-chip-eyeq4-visual-processing-ad-as-and-automated-driving-design>. [Hozzáférés dátuma: 04 09 2019].
- [52] MobileEye, „The Evolution of EyeQ,” MobileEye, 2018. [Online]. Available: <https://www.mobileye.com/our-technology/evolution-eyeq-chip/>. [Hozzáférés dátuma: 08 09 2019].
- [53] Renesas, „Renesas R-Car V3H SoC,” Renesas, 24 10 2018. [Online]. Available: <https://www.renesas.com/br/en/solutions/automotive/soc/r-car-v3h.html>. [Hozzáférés dátuma: 04 09 2019].
- [54] J. Yoshida, „Renesas SoC to Double EyeQ5 Efficiency,” *EE/Times*, p. https://www.eetimes.com/document.asp?doc_id=1332924, 02 02 2018.
- [55] G. d. Mestral, „Velcro”. Svájci Szabadalom száma: US3009235A, 1955.

- [56] Y. Xiao, *Bio-Inspired Computing and Networking*, USA: CRC Press, 2011.
- [57] G. B. Edwin, *Visual Perception of Depth and Distance*, New York: Appleton Century Crofts, 1942, pp. 263-311..
- [58] G. Laurent és G. Gabbiani, „Collision-avoidance: nature's many solutions,” *Nature neuroscience*, Vol. 1., pp. 261-263, 1998.
- [59] M. O'Shea, C. Rowell, H. F és J. L. D. Williams, „The anatomy of a locust visual interneurone: the descending contralateral movement detector,” *Journal of experimental Biology*, Vol. 60., pp. 1-12, 1974.
- [60] H. Sun és B. J. Frost, „Computation of different optical variables of looming objects in pigeon nucleus rotundus neurons,” *Nature neuroscience*, Vol. 1., pp. 296-303, 1998.
- [61] W. Schiff, J. A. Caviness és J. J. Gibson, „Persistent Fear Responses in Rhesus Monkeys to the Optical Stimulus of "Looming",” *Science*, Vol. 136., pp. 982-983, 1962.
- [62] M. Wicklein és N. Strausfeld, „Organization and significance of neurons that detect change of visual depth in the hawk moth *Manduca sexta*,” *The Journal of Comparative Neurology*, Vol. 424., pp. 356-376., 2000.
- [63] P. G. Jabłoński és N. J. Strausfeld, „Exploitation of an ancient escape circuit by an avian predator: prey sensitivity to model predator display in the field,” *Brain, Behaviour and Evolution*, Vol. 56., pp. 94-106., 2000.
- [64] W. Yongchang és J. F. Barrie, „Time to collision is signalled by neurons in the nucleus rotundus of pigeons,” *Nature*, Vol. 356., pp. 236-238., 1992..
- [65] W. Ball és E. Tronick, „Infant responses to impending collision: optical and real,” *Science*, Vol. 171, pp. 818-820., 1971.
- [66] R. D. Hamer, „What can my baby see?,” *Parents' Press*, Vol. 11., 1990.
- [67] R. S. Allison, B. J. Gillam és E. Vecellio, „Binocular depth discrimination and estimation beyond interaction space,” *Journal of Vision*, Vol. 9., pp. 1-14, 2009.

- [68] W. Schiff, „Perception of impending collision: A study of visually directed avoidant behavior,” *Psychological Monographs: General and Applied*, Vol. 79., pp. 1-26., 1965.
- [69] A. v. Mühlénen és A. Lleras, „No-onset looming motion guides spatial attention,” *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 33., pp. 1297-1310, 2007.
- [70] C. Cappea, G. Thutb, V. Romeib és M. M. Murray, „Selective integration of auditory-visual looming cues by humans,” *Neuropsychologia*, Vol. 47., pp. 1045-1052, 2009.
- [71] I. Gat, M. Benady és A. Shasua, „A Monocular Vision Advance Warning System for the Automotive Aftermarket,” *SAE Technical Paper*, 2005.
- [72] Mobileye, „About Us - Mobileye,” Mobileye, 2018. [Online]. Available: <https://www.mobileye.com/about/>. [Hozzáférés dátuma: 17. március 2018.].
- [73] W.-C. Chang és C. Chih-Wei, „Online Boosting for Vehicle Detection,” *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS*, Vol. 40, pp. 892-902, 2010.
- [74] A. Mukhtar, L. Xia és T. B. Tang, „Vehicle Detection Techniques for Collision Avoidance Systems: A Review,” *IEEE Transaction on Intelligent Transportation Systems*, Vol. 16, pp. 2318-2338, 2015.
- [75] G. Liñán-Cembrano, L. Carranza, C. Rind, Á. Zarandy, M. Soininen és A. Rodríguez-Vázquez, „Insect-vision inspired collision warning vision processor for automobiles,” *IEEE CIRCUITS AND SYSTEMS MAGAZINE*, Vol. 8., pp. 6-24., 2008.
- [76] R. Stafford, R. D. Santer és F. C. Rind, „A bio-inspired visual collision detection mechanism for cars: Combining insect inspired neurons to create a robust system,” *BioSystems*, Vol. 84., pp. 164-171, 2007.
- [77] R. Hartley és A. Zisserman, *Multiple View Geometry in computer vision*, 2nd edition, Cambridge: Cambridge, 2003.
- [78] A. Z. Richard Hartley, *Multiple View Geometry in computer vision*, Cambridge, 2003 (2. kiadás).

- [79] A. T. Lin, „Visual collision detector for a mobile robot,” *Mobile Robotic Lab McGill University*, 1995.
- [80] Stanford University, „Elements of Computer Vision,” Stanford University, 1998. [Online]. Available: <https://cs.stanford.edu/people/eroberts/courses/soco/projects/1997-98/computer-vision/motion.html>. [Hozzáférés dátuma: 15 05 2018].
- [81] M. Kiricarslan és Z. Y. Yu, „Predict Vehicle Collision by TTC From Motion Using a Single Video Camera,” *IEEE Transactions on Intelligent Transportation Systems*, pp. 1-12, 2018.
- [82] F. Yang, W. Choi és Y. Park, „Exploit All the Layers: Fast and Accurate CNN Object Detector with ScaleDependent Pooling and Cascaded Rejection Classifiers,” in *IEEE Conference on Computer Vision and Pattern Recognition Conference*, Las Vegas, 2016.
- [83] L. Carranza, R. Laviana, S. Vargas, J. Cuadri, G. Linan, E. Roca és A. Rodriguez-Vazquez, „Locust-inspired vision system on chip architecture for collision detection in automotive applications,” *IEEE International Symposium on Circuits and Systems*, 2006.
- [84] T. Kohler, F. Rochter, J. Lindemann és R. Moller, „Bio-inspired motion detection in an FPGA-based smart camera module,” *Bioinspiration & Biomimetics*, Vol. 4., 2009.
- [85] A. C. Damon és B. D. Jonathan, „Parallel Computations in Insect and Mammalian Visual Motion Processing,” *Processing. Current Biology*, Vol. 26., pp. R1062-R1072, 2016.
- [86] J. M. Ache, J. Polsky, S. Alghailani, D. D. Bock, C. R. v. Reyn és G. M. Card, „Neural Basis for Looming Size and Velocity Encoding in the Drosophila Giant Fiber Escape Pathway,” *Current Biology*, Vol. 29., pp. 1073-1081.
- [87] L. Mu, J. P. Bacon, K. Ito és N. J. Strausfeld, „Responses of Drosophila giant descending neurons to visual and mechanical stimuli,” *Journal of Experimental Biology*, Vol. 217., pp. 2121-2129, 2014.

- [88] R. D. Santer, R. Stafford és F. C. Rind, „Retinally-generated saccadic suppression of a locust looming-detector neuron: investigations using a robot locust,” *J. Royal Society Interface, Vol. 1.*, pp. 61-77, 2004.
- [89] J. Cuadri, G. Liñan-Cembrano, E. Roca és A. Rodríguez-Vázquez, „Macromodeling for analog design and robustness boosting in bio-inspired computing models,” *Proceedings of SPIE - The International Society for Optical Engineering*, pp. 1-12, 2005.
- [90] G. Liñán-Cembrano, L. Carranza, C. Rind, A. Zarandy, M. Soininen és A. Rodríguez-Vázquez, „Insect-Vision Inspired Collision Warning Vision Processor for Automobiles,” *IEEE CIRCUITS AND SYSTEMS MAGAZINE*, Vol. 8., pp. 6-24, 2008.
- [91] F. C. Rind, „Non-directional, movement sensitive neurones of the locust optic lobe,” *Journal of Comparative Physiology*, Vol. 161, pp. 477-494, 1987.
- [92] P. J. Simmons és F. C. Rind, „Responses to object approach by a wide field visual neurone, the LGMD2 of the locust: Characterization and image cues,” *Journal of Comparative Physiology*, Vol. 180, pp. 203-214, 1997.
- [93] Q. Fu és S. Yue, „Modelling LGMD2 Visual Neuron System,” in *IEEE International Workshop on Machine Learning for Signal Processing*, Boston, 2015.
- [94] Q. Fu, S. Yue és C. Hu, „Bio-inspired Collision Detector with Enhanced Selectivity for Ground Robotic Vision System,” in *British Machine Vision Conference*, Lincoln, 2016.
- [95] W. Yongchang, S. Jiang és J. F. Barrie, „Visual processing in pigeon nucleus rotundus: luminance, color, motion, and looming subdivisions,” *Visual Neuroscience, Vol. 10.*, pp. 21-30, 1993.
- [96] H. S. Barrie J. Frost, „The Biological Bases of Time-to-collision Computation,” in *Advances in Psychology*, Amsterdam, 2004, pp. 13-37.
- [97] W. Ling, S. Hongjing és Y. Dezhong, „A Neural Network Model for the Estimation of Time-to-Collision,” in *Advances in Neural Networks - ISNN 2006*, Berlin, Springer, 2006, pp. 614-619.

- [98] L.-Q. Wu, Y.-Q. Niu, J. Yang és S.-R. Wang, „Tectal neurons signal impending collision of looming,” *European Journal of Neuroscience*, Vol. 22, pp. 2325-2331, 2005.
- [99] J. R. Gray, E. Bincow és R. M. Robertson, „A pair of motion-sensitive neurons in the locust encode approaches of a looming object,” *Journal of Comparative Physiology*, Vol. 196, pp. 927-938, 2010.
- [100] F. C. Rind és P. J. Simmons, „Seeing what is coming: building collision-sensitive neurones,” *Trends in Neuroscience*, Vol. 22., pp. 215-220., 1999.
- [101] Z. Szegedi, Ragadozómadarak Abaújban és Zemplénben, Abaúj-Zemplén Értékeiért Közhasznú Egyesület, 2007.
- [102] Á. Zarándy és T. Fülöp, „Retinal approaching object detector model implementation and validation,” in *IEEE international symposium on circuits and systems - ISCAS*, Paris, 2010.
- [103] Á. Zarándy és T. Fülöp, „Implementation and Validation of a Looming Object Detector Model Derived from Mammalian Retinal Circuit,” in *Focal-Plane Sensor-Processor Chips*, Springer, 2011, pp. 245-260..
- [104] T. Fülöp és Á. Zarándy, „Bio-inspired looming direction detection method,” in *13th International Workshop on Cellular Nanoscale Networks and their Applications - CNNA*, Turin, 2012.
- [105] T. Fülöp és Á. Zarándy, „Bio-inspired looming object detector algorithm on the Eye-RIS focal plane-processor system,” in *12th international workshop on cellular nanoscale networks and their applications - CNNA*, Berkeley, 2010.
- [106] M. Habibi, „Application of dynamic counter circuits in programmable digital pixel sensor architectures,” *International Journal of Circuit Theory and Applications*, Vol. 42., pp. 989-1005, 2014.
- [107] S. Han és J. Kim, „Video Scene Change Detection Using Convolution Neural Network,” in *International Conference on Information Technology*, Singapore, 2017.
- [108] C.-S. Bouganis és M. Brookes, „Statistical multiple light source detection,” *IET Computer Vision*, Vol. 1, pp. 79-91, 2007.

- [109] J.-H. Park és C.-s. Jeong, „Real-time Signal Light Detection,” *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 2008.
- [110] G. Yu, A. Lei, H. Li, Y. Wang, Z. Wang és C. Hu, „A Real-Time Traffic Light Detection Algorithm,” *SAE International by Univ of Nottingham*, pp. 1-10, 2018.
- [111] Z. Wu, Y. Watanabe és M. Ishikawa, „Hybrid LED Traffic Light Detection Using High-speed Camera,” in *IEEE 19th International Conference on Intelligent Transportation Systems*, Rio de Janeiro, 2016.
- [112] C. E. Shannon, „Communication in the presence of noise,” *Proceedings of the Institute of Radio Engineers*, Vol. 37., pp. 10-21., 1949.
- [113] P. Aradi, J. Graff és G. Lipovszki, Számítógépes szimuláció, Budapest: BME MOGI, 2014.
- [114] X. Huang, R. Bai, X. Jin és H. Fu, „Robust and Efficient Frequency Estimator for Undersampled Waveforms Based on Frequency Offset Recognition,” *PLOS ONE*, pp. 1-12., 2016.
- [115] G. Goertzel, „An Algorithm for the Evaluation of Finite Trigonometric Series,” *The American Mathematical Monthly*, Vol. 65., pp. 34-35., 1958.
- [116] J.-Y. Tinevez, *SIMPLETRACKER a simple particle tracking algorithm that can deal with gaps*, Jean-Yves Tinevez, 2012.
- [117] T. Fülöp és Á. Zarándy, „Chromatic Spatial-temporal Frequency Analysis for Traffic Light detection,” in *4th IEEE International Conference on Cognitive Infocommunications - CogInfoCom*, Budapest, 2013.
- [118] Á. Zarándy és T. Fülöp, „Approaching object detector mouse retina circuit model analysis and implementation on cellular sensor-processor array,” *INTERNATIONAL JOURNAL OF CIRCUIT THEORY AND APPLICATIONS*, Vol. 40., pp. 1249-1264., 2012.
- [119] Nayuki, „How to implement the discrete Fourier transform,” 06. 06. 2017.. [Online]. Available: <https://www.nayuki.io/page/how-to-implement-the-discrete-fourier-transform>. [Hozzáférés dátuma: 01. 05. 2018.].
- [120] L. Trammell, „Detecting A Single Frequency Efficiently,” Microstar Laboratories, Inc., 11. 2005.. [Online]. Available:

<http://www.mstarlabs.com/dsp/goertzel/goertzel.html>. [Hozzáférés dátuma: 01. 05. 2018.].

[121] A. Horváth, *CNN dynamics*. [Performance]. PPKE-ITK, 2019.

