

Pozíció-meghatározás sztereo kamerarendszer segítségével

***Reizinger Patrik* *(W5PDBR)***

**Konzulens:** Kertész Zsolt

2018.02.13.

Tartalom

[Problémamegfogalmazás 3](#_Toc513413241)

[**Szakirodalmi áttekintés** 3](#_Toc513413242)

[Optical flow 3](#_Toc513413243)

[Mozgás alapú sztereo (motion stereo) 4](#_Toc513413244)

[Sztereo kamerarekonstrukció két kamerával 5](#_Toc513413245)

[Sztereo rekonstrukciós algoritmus 5](#_Toc513413246)

[Kamerakalibráció: 5](#_Toc513413247)

[Epipoláris geometria: 7](#_Toc513413248)

[Rektifikáció: 8](#_Toc513413249)

[Reprojekció: 9](#_Toc513413250)

[Implementáció: 10](#_Toc513413251)

[Hardveres megvalósítás: 10](#_Toc513413252)

[Szoftveres kialakítás: 11](#_Toc513413253)

[Kalibráció: 11](#_Toc513413254)

[Futtatás: 12](#_Toc513413255)

[Értékelés, fejlesztési lehetőségek: 12](#_Toc513413256)

# Problémamegfogalmazás

Különböző képfeldolgozási feladatok során gyakran kap központi szerepet a kamerakép(ek) alapján történő háromdimenziós térrekonstrukció, legyen szó robotikára, önvezető autókról, drónokról vagy akár méréstechnikai feladatokról.

A jelen dolgozatban tárgyalt megoldás egy virtuálisvalóság-rendszerhez fejlesztett kontroller relatív pozíciószabályozó algoritmusának ellenőrzése céljából került megvalósításra, melynek segítségével a perféria szenzorainak pontossága egyszerűen mérhető, s így lehetőség nyílik a korrekcióra.

A megfogalmazott probléma abból eredeztethető, hogy a képérzékelőre történő leképezés, a perspektív transzformáció során a mélységi dimenzió elveszítjük, lévén a kamera képrézékelője kétdimenziós érzékelést tesz lehetővé, így szoftveres megoldásokra van szükség ahhoz, hogy a rendelkezésre álló információk segítségével a mélységi információ visszanyerhető legyen.

# **Szakirodalmi áttekintés**

Háromdimenziós rekonstrukciós feladatokra számos megközelítésmód, illetve algoritmus létezik, melyek mind hardverigényüket, mind pedig előnyeiket, ill. szűk keresztmetszetüket tekintve különbözőek, így a tervezési fázis során kiemelten fontos végig gondolni az optimalizálni kívánt metrikát.

Abból kifolyólag, hogy az egyes módszerek (legalább részben) különböző szakirodalmi háttérre támaszkodnak, így először a módszer kiválasztására kerül sor, majd ezután annak részletes elemzésére.

## Optical flow

Az optical flow, másnéven optikai áramlás módszere segítségével egyetlen kamerával is végrehajtható a háromdimenziós térrekonstrukció. Ahhoz, hogy a módszer magját alkotó intenzitás-áramlás egyenlet, vagyis

ahol a *t* index az idő, míg *x,y* indexek a térbeli deriváltat jelölik, teljesüljön, három korlátozó feltételnek is teljesülnie kell:

* Egy adott pont szürkeségi szintjét időben (közel) állandónak kell lennie
* A követni kívánt objektum sebessége, vagyis *dx, ill. dy* abszolútértéke kellően kicsi – annak érdekében, hogy az intenzitás-áramlási egyenlet jelen formában felírható legyen, az ugyanis egy Taylor-soros közelítés magasabb rendű tagjainak elhagyásával fejezhető ki a fenti alakban, ennek elérése pedig csak kis értékű deriváltak esetében lehetséges
* Az azonos objektumokhoz tartozó pontok elmozdulása (közel) azonos

Azonban ezen feltételek kielégítése mellett sem egyértelmű a probléma megoldása., ugyanis az intenzitás-áramlás egyenlet aluldeterminált, a sebességgradiens két ismeretlen rendezője nem fejezhető ki egyértelműen egy egyenlet segítségével.

Habár a Lucas-Kanade-eljárás (illetve annak iteratív változata) megoldást jelent az aluldeterminált egyenlet megoldására, továbbra sem garantálja a három korlátozó feltétel teljesülését – így bár az optical flow több esetben kielégíthető megoldást jelenthet, azonban a többkomponensű algoritmikus alrendszer, amelyre szükség van a kielégítő, robusztus megoldás megtalálásához, nem teszi alkalmassá a vázolt feladat ellátására.

## Mozgás alapú sztereo (motion stereo)

Rekonstrukciós eljárások esetén felhasználható annak ténye is, hogy amennyiben egy adott kamera, ill. a vizsgált objektum között relatív elmozdulás figyelhető meg, akkor az objektum képsíkra vetített mérete megváltozik, ami alapján következtethetünk az objektum távolságára.

A motion stereo algoritmusa, habár felettébb egyszerű, mivel hasonló háromszögek segítségével történik a mélységi információ nyomon követése, azonban mivel jelen esetben kis elmozdulások követésére van szükség, numerikus instabilitást okozhat a kis elmozdulás-értékekkel történő leosztás, ugyanis a Z koordináta a következő formula szerint számítható:

ahol *b* az ún. bázistávolság (mozgó kamera esetén a kamera elmozdulása), *r* a képsík optikai középponttól mért távolsága, míg *D* a diszparitás, vagyis a képpont képének két időpillanatban vett vetületének euklideszi távolsága.

A numerikus instabilitás miatt további problémát vet fel, ha az eljárást nem hagyományos módon (ti. a kamerát mozgatjuk) szeretnénk használni, ugyanis az objektum elmozdulása esetén a *b* paraméter ismeretlen, így alulhatározott lesz a formula.

## Sztereo kamerarekonstrukció két kamerával

Az eddig bemutatott módszerekkel ellentétben a klasszikus sztereo alkalmazások több kamera képét, ill. a képek közötti eltérést, a diszparitást használják fel a mélységi információ visszanyerésére.

Ezek az eljárások általában nagyobb számításikapacitás-igénnyel rendelkeznek, azonban ennek következtében pontosabb, robusztusabb megoldást eredményező eljárásokkal is kiegészítésre került, így a rendelkezésre álló erőforrásoknak megfelelően meghatározható a kompromisszum pontosság-erőforrásigény dimenziókat tekintve.

A következő szakaszban a sztereo kamerarekonstrukciós algoritmus lépéseinek részletes bemutatására kerül majd sor.

# Sztereo rekonstrukciós algoritmus

A mélyésginformáció kinyerése érdekében egy több lépésből felépített eljárásra van szükségünk, melnyek során a következő fázisok megvalósítására van szükség:

* Kamerakalibráció: a kamera torzításának kompenzálása
* Epipoláris geometria meghatározása: a két kamera között fennáló geometriai transzformáció kiszámítása
* Rektifikáció: a két kamerakép összetartozó pontjai megkeresését megkönnyítendő
* Sztereo megfeleltetés: a kamerák képeinek egymáshoz történő rendelése, a diszparitás kiszámítása
* Reprojekció: a diszparitás által meghatározott mélyésginformáció segítségével a képsíkra projektált képpontok valós objektumpontokká történő transzformálása

## Kamerakalibráció:

Az olcsó CMOS-szenzorokkal felszerelt kamerák elterjedésével egyidejűleg széles körben jelentkező probléma a kameraberendezések nemideális voltának, torzításainak korrekciója.

Alapvetően kétféle torzítási jelenséget különböztetünk meg, melyek mind jellegüket, mind pedig eredetüket tekintve jól elkülöníthetőek.

A radiális torzítás a lencse geometriájából adódó nemidealitás, ugyanis a matematikailag ideális parabolikus lencsékkel szemben a gyártási folyamatok során jóval egyszerűbben előállítható szferikus lencsék a lencse pereméhez közeli pixeleket jelentős mértékben torzíthatják.

A radiális torzítás megjelenési formája szintén kétféle, egyrészt lehet ún. hordótorzítás, ami egy négyzetet oly módon képez le, hogy annak egyik szimmetriatengelyénél megnyújtja a képet, míg a párnatorzítás esetén éppen fordított jelenségnek lehetünk tanúi, vagyis a középső tartomány összenyomásának.

Annak köszönhetően, hogy a radiális torzítás a képközéppontban 0, így jól közelíthető az origó körüli Taylor-sorral – itt ráadásul, figyelembe véve a szimmetriaviszonyokat, vagyis, hogy a torzítás szimmetrikus az origótól mért *r* távolságra, a sugárra, csak a páros hatványú tagokat kell figyelembe vennünk.

Ezek alapján a radiális torzítás korrekciós formulája a következő alakban írható fel:

ahol a 0 index jelöli az eredeti, míg a *rad* index a radiális torzítástól korrigált képpontokat.

A tangenciális torzítás alapvetően a nem megfelelő minőségbiztosítási paramétereknek köszönhető, ugyanis ilyen jellegű torzítás tipikus kiváltó tényezője a nem megfelelően (pl. nem a lencse optikai tengelyére merőlegesen) rögzített képérzékelő szenzor.

A tangenciális torzítás két további taggal bővíti a torzítási tényezők paramétervektorát, így összességében 5 paraméter kiszámítására van pusztán a lencse, ill. a képrézékelő nemidealitása miatt.

## Epipoláris geometria:

A kameraparaméterek korrekcióját követően szükség van a két kamera geometriáját összerendelő transzformáció meghatározására, melynek leírására az epipoláris geometria nyújtotta lehetőségek kerültek alkalmazásra.

Epipoláris geometria alatt a két kamera optikai középpontjai, illetve a leképezendő objektumpont által meghatározott geometriát értjük, mely három kitüntetett típusú komponenssel rendelkezik:

* Epipoláris sík: az optikai középpontok, valamint az objektumpint által meghatározott sík
* Epipoláris egyenes: a perpekstív transzformáció következtében az egyik képérzékelőre leképezett objektumpont rekonstrukciója (annak képéből) a másik képérzékelőn egy egyenesként jelenik megy a mélységi információ bizonytalanságából kifolyólag
* Epipólus: adott optikai középpont másik képérzékelőre vetített képe

Nyilvánvaló, hogy a két kamera között található olyan transzformáció, amely az egyik kamera képét a másik kamera képébe transzformálja, ennek meghatározásához a rendelkezésre álló információktól függően több lehetőség adódik:

* Amennyiben a két kamera relatív helyzete, valamint a belső paraméterek (fókusztávolság, képközéppont) ismertek, akkor meghatározhatók az epipoláris egyenesek
* Amennyiben csak a belső (ún. intrinsic) paraméterek ismertek, akkor szükséges az ún extrinsic paraméterek meghatározása is, amely a két kamerakép közötti átjárást határozza meg (ez *essential matrix* néven ismert a szakirodalomban). Az essential matrix egy *R* forgatási, illetve egy *S* transzlációs mátrix szorzataként áll elő
* Mind az intrinsic, mind az extrinsic paraméterek hiányában olyan transzformáció meghatározására van szükségünk, amely a világkoordináták és a pixelkoordináták között teremt kapcsolatot (köztes lépésként értelemszerűen a perspektív transzformációt is magába foglalva) – az ezen leképezést leíró mátrix *fundamental matrix* néven szerepel a szakirodalomban. A fundamental matrix tartalmazza az intrinsic paramétereket, így a fókusztávolságokat, ill. a képközéppontok koordinátáit.

## Rektifikáció:

A rektifikáció művelete az általános sztereo kamerapár, illetve a párhuzamos optikai tengelyű sztereo kamerapár közötti transzformációt írja le.

A rektifikáció előnye, hogy az összetartozó objektumpont-epipoláris egyenes párok összerendelése során a egy szabadsági fok eliminálható, a rektifikált képeken ugyanis az epiloáris egyenesek vízszintessé válnak, vagyis a kétdimenziós megfeleltetési probléma egydimenzióssá egyszerűsíthető, ami számításikapacitás tekintetében jelentős megtakarítás – még azon megfontolások mellett is, hogy a rektifikációt minden képkockára végre kell hajtani, azonban a transzformáció leírására, a rektifikációs paraméterek meghatározására csak egyszer van szükség.

Ez az alábbi gondolatmenetet követve egyszerűen belátható: amennyiben adott egy dimenziós képkocka, akkor *worst case* esetben egy pixelhez összerendelési műveletet kell elvégezni (ahol , jelöli a képsorok számát), melynek számításikapacitás-igénye elemszámú konvolúciós ablak esetében , míg a rektifikáció során ez csökken, ugyanis a rektifikáció eredménye egy lookup tábla (LUT), melynek segítségével könnyen elvégezhető a leképezés (amennyiben figyelembe vesszük az alkalmazott interpolációt is, akkor is nagyságrendileg kevesebb művelettel elvégezhető a megfeleltetés).

A transzformáció meghatározása történhet Hartley vagy Bouguet módszere alapján: míg előbbi előnye, hogy 7 képpár alapján már képes a rektifikációra, azonban az eljárás nem képes abszolút skála meghatározására, vagyis a közeli, de kis méretű, valamint a nagy méretű távoli objektumok között ez alapján nem lehet különbséget tenni.

A Hartley-algoritmus esetében ezt a problémát sikerül megoldania, az optimalizálandó célfüggvény ebben az esetben a reprojekciós hiba (minimum), illetve a közös képterület (maximum) – természetesen mindezt nagyobb számításikapacitás mellett képes csak megvalósítani, tehát a feladatnak megfelelően kell algoritmust választani.

Eddig a pontig bezárólag meghatározásra került a diszparitás-térkép, ahhoz azonban, hogy megfelelő minőségű háromdiemnziós rekonstrukciót tudjunk megvalósítani, további finomhangolásra van szükségünk, ugyanis a számított struktúra az esetek döntő többségében egy ritka, ráadásul zajjal terhelt mátrix. Ez többnyire abban nyilvánul meg, hogy a diszparitás-térkép jelentős része „lyukakkal” borított, vagyis érvénytelen adatot tartalmaz. Az ilyen jellegű probléma egy simító jellegű szűrővel, például a súlyozott legkisebb négyzetek (WLS, weighted least squares) módszerével kiküszöbölhető.

A súlyozott legkisebb négyzetek módszere gyakran alkalmazott eljárás alacsony szintű képfeldolgozásban, alapvetően az intuíció mögötte energiaalapú, vagyis adott tartomány energiaszintjének minimalizására törekszik, tehát

a W paramétervektort keressük a β helyfüggő súlyozóvektor jelenléte mellett, ami lehetővé teszi a lyukak betöltését a ritka mátrixban.

## Reprojekció:

A sztereo megfeleltetés során meghatározott lookup táblával minden képkockára könnyedén számítható a diszparitás, azonban az így kapott ponthármasokat viossza kell transzformálni a világkoordinátarendszerbe, amely egy újabb transzformációs mátrix, az ún. *reprojekciós mátrix (Q)* segítégével tehető meg – Q pl. a jelen feladatban alkalmazott Bouguet-algoritmus során előáll, így az implementáció során ennek kiszámítására külön nem kell figyelmet fordítani.

Q tehát tartalmazza egyrészt a bal kamera intrinsic paramétereit, továbbá a jobb oldali kamera képközéppontjának x koordinátáját a jobb oldali kamera koordináta-rendszerében (cx’, amely rektifikált sztereo kamerapár estében megegyezik cx értékével), illetőleg a kamerák közötti transzlációs vektor x koordinátáját (Tx), ami a két kamera optikai tengelye közötti távolság.

Q ismeretében a ponthármas egy mátrixszorzás segítségével transzformálható vissza a világkoordinátarendszerbe.

Amely formulában *d* az *x, y* pontpárhoz tartozó diszparitásértéket jelöli. Az így kapott homogén koordinátarendszerből három dimenzióba a homogén koordinátarendszerek konvenciójának megfelelően, vagyis a *W* koordinátával történő, koordinátánkénti leosztással transzformálhatjuk az adott pontot, vagyis

lesz az adott pont koordinátája.

# Implementáció:

Jelen fejezet a hardveres, illetve a szoftveres megvalósítás részleteit mutatja be, kitérve az adott feladat specifikus tényezőire.

## Hardveres megvalósítás:

Az első alfejezetben a rendszer hardveres sajátosságai kerülnek ismertetésre. A feladat során használt sztereo kamerapár a Logitech C270-es típusszámú, 720x480 pixeles felbontású modellje volt.

Ez a webkameratípus fixfókuszos kialakítású, csatolófelülete USB2.0, így reálisan 30 FPS (*frame per second*, képkocka/másodperc) az elvárható adatátviteli sebesség.

A két webkamera egy háromlábú kameraállványra került rögzítésre, az ehhez szükséges kerszttartót 3D-nyomtatóval készítettem, a mechanikai terveket pedig az Autodesk cég Inventor 2018 programjában terveztem meg. A kialakítás alapvetően párhuzamos tengelyű kamerapárhoz lett tervezve, természetesen azonban az emberi szem modelljének megfelelő, néhány fokos párhuzamos tengelytől való eltérés is megvalósítható.

## Szoftveres kialakítás:

A korábbiakban ismertetett algoritmusokat Python nyelven implementáltam az OpenCV nyílt forráskódú, elsődlegesen gépi látásra, képfeldolgozásra optimalizált függvénykönyvtár alkalmazásával.

A program alapvetően a parancssoron keresztül biztosít interfészt a felhasználó számára, a következő beállítások érhetők el:

* Kalibráció indítása
* Biztonsági kapcsoló az újrakalibrációhoz, így véletlenül nem törölhetők a korábbi kalibráció eredményei
* A tesztábrák számának meghatározása kalibrációhoz
* Kapcsolók a futtatás során a különböző fázisok (rektifikált kép, diszparitástérkép, stb. megjelenítéséhez)

## Kalibráció:

A kalibrációs folymat alapját az OpenCV beépített, kifejezetten kalibrációs célokra implementált függvénye alkotja, amely képes adott méretű sakktáblamintázat felismerésére, a mintázat fizikai méreteinek megadása által pedig a kalibráció elvégezhető.

A kalibráció folyamata tehát felvételek készítése a kalibrációs mintázatról sikeres mintázatfelsimerési peremfeltétel mellett. Kalibrációhoz azonban a kamerák által biztosított, pixel szerinti felbontás nem feltétlen elegenedő – azonban az OpenCV rendelkezik a problémát megoldani képes, szubpixelees mintázat (pontosabban sarokpont, ez azonban a sakktábla esetében pontosan megegyezik a mintázatot jellemző karakterisztikus pontok halmazával) meghatározó algoritmussal, amely egy megválasztott kezdeti érték környezetében kiválasztott pontokból alkot egyenletrednszert

alakban, amely egy P pontra vonatkoztatott szürkeségi szintet jellemző gradiens, valamint a P és Q pontok közötti vektor skalárszorzata. Az egyenletről azért tehető fel, hogy a jobb oldal azonosan 0, vagyis hogy a gradiens és a két pontot összekötő szakasz merőlegesek, mivel ha P belső pont (egy homogén intenzitású tartományt tekintve), akkor az intenzitásgradiens 0, amennyiben pedig tartományhatárra, vagyis élre esik, akkor a P-Q vektor az éllel párhuzamos, míg a gradiens arra merőleges.

A rögzített képek alapján először külön-külön kameránként, majd ennek eredméynét a sztereo kalibráció kiindulási éretékeként felhasználva történik meg a szükséges kameramátrixok kiszámítása, majd végüla rektifikációhoz szükséges lookup tábla, ill. a Q reprojekciós mátrix meghatározása, melyek a művelet végeredményeként hdf5 formátumban mentésre kerülnek, hogy a későbbiek folyamán a futtatást megelőzően ne kelljen minden alkalommal végrehajtani a kalibrációt.

## Futtatás:

A mervlemezre mentett kalibrációs paramétereket beolvasva, majd a rektifikációt elvégezve lépünk be a futtatás során megvalósított hurokba, melynek további feladata adott színű marker nyomonkövetése.

Ehhez az RGB képet HSL színtérbe kovertálom, mivel így adott színre nagyobb megbízhatósággal lehet szűrni, révén az árnyalat, színtelítettség, megvilágítottság paraméterek közvetlen kapcsolatban állnak az adott színnel, nem pedig valamilyen lineárkombinációban, mint az RGB színtér esetén.

A megfelelő szűrési patraméterek eredményeként visszakapott bináris maszk további finomítására morfológiai műveletek segítségével kerül sor, elsősorban a kisebb, zajnak tekinthető kisméretű objektumok eliminálására erózióval, majd lyukak kitöltésére zárással.

A feldolgozott maszkon belül az OpenCV beépített kontúrkereső rutinja segítségével a legnagyobb területű objektumon kívül az összes többi kiszűrésre kerül, az így megkapott markerobjektum köré téglalapot illesztve, majd annak középpontját kiszámítva egyetlen pontpárral jellemezni tudjuk a markert, illetve ezen pontkoordinátákkal a diszparitástérképbe indexelve vissza tudjuk állítani a marker háromdimenziós pozícióját – ezen lépésnél természetesen érdemes valamilyen régióalapú kijelölést használni, és a mélységi információt szűrni.

# Értékelés, fejlesztési lehetőségek:

A kialakított, sztereo kamerapáron alapuló háromdimenziós rekonstrukciós keretrendszer az kalibrációs metrikát tekintve remek eredményekkel rendelkezik, az egyes kamerákra vonatkoztatott projekciós hiba (vagyis a kalibráció révén előálló mátrixok által a képsíkra transzformált, világkoordinátákkal rendelkező pontok és a kamersíkon található képpontok átlagos négyzetes hibája) megfelelő számú (jelen esetben 15 darab) és minőségű kalibrációs ábra mellett 0.4 pixel alatti, ami a sztereo kalibráció hibájára is teljesülő felső korlát.

X, ill. Y irányban a módszer centiméter alatti pontossággal távolságtartó, a mélységi információt tekintve azonban, a képkockák közötti temporális összefüggések felhasználásának hiányában, illetőleg a diszparitástérképen a marker középpontjával történő indexeléssel ez a pontosság nem érhető el, így a továbbiakban ezen a téren cél a minél kifinomultabb megoldás megvalósítása.