

Képfeldolgozás

Stéger József

Szüle János

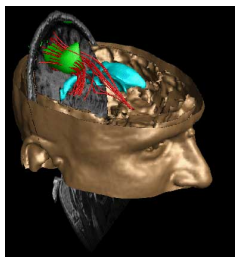
2017. szeptember 25.

1. Bevezető

Felgyorsult és folyamatosan fejlődő világunkban elengedhetetlen, hogy a képi információkat gyorsan és pontosan fel tudjuk dolgozni. Az első számítógépek megjelenésekor, azok számítási kapacitása még nem tette ezt lehetővé, manapság viszont egy jobb minőségű mobiltelefonnal is képesek lehetünk akár mozgó képek valós időben történő feldolgozására.

Mind a statikus, mind a dinamikus képfeldolgozásnak számos alkalmazási területe ismert, lásd ízelítőként az 1. ábra példáit illetve a következő felsorolást.

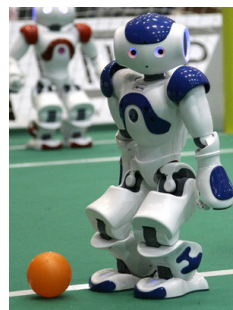
- A robotika (3D képkötés, tárgyfelismerés és -mozgatás),
- az orvostudomány (CT, MRI, PET),
- a filmipar (mozgás detektálás és *motion capture*),
- az objektumfelismerés (vonalkódolvasás, arc-, betűfelismerés),
- a megfigyelő és biztonsági rendszerek.



(a) MRI head



(b) Motion capture



(c) Robotika

1. ábra. Példák a képfeldolgozás szerteágazó témaköréből.

Természetesen ez a lista messze nem teljes, számos érdekes és fontos ponttal lehetne bővíteni. Valójában attól a ponttól kezdve, hogy a minket körül ölelő világból érkező jeleket kinyerjük és az adatokat digitális formába öntjük, a felhasználási lehetőségeknek csak a képzelet szab határt.

A képfeldolgozáshoz köthető mérésekhez és hiteles adatfeldolgozáshoz általában szükségünk van arra, hogy pontosan ismerjük a teljes képalkotó rendszer fizikai paramétereit (pl.: fókusz, mérési normális, stb.). Ugyanis ezek a paraméterek teremtik meg a kapcsolatot a mérendő világ és a képi tér között. A képalkotó rendszerek vetületeket készítenek terünkről, és feladat ezen vetületek értelmezése esetlegesen több forrásból érkező vetületinformációk illesztése, hogy rekonstruálni tudjuk az eredeti térben értelmezett pontokat, távolságokat. Az ilyen jellegű feladatokban közös, hogy alulhatározott problémának kell tekinteni őket. Tovább nehezíti a feladatot, ha a képi információ zajjal terhelt. A zaj tipikus forrásai lehetnek a visszaverődések, a lencsehibákból eredő fókusz sík torzulás valamint a képalkotó rendszer tökéletlen függesztéséből eredő rezgések, mozgások. Vannak olyan problémák, ahol figyelembe kell venni, hogy a mérés során a vizsgált tárgy helyzete vagy alakja megváltozik, és ezeket a változásokat nyomon kell kísérni és modellezni kell egy-egy képsorozatnál. Gyakran nehéz minden tekintetben tökéletesen modellezni a tárgy változó viselkedését. Az ilyen problémát egy részét a mesterséges intelligencia és a képfeldolgozás ötvöztetésével próbálják megoldani.

A képfeldolgozási problémák három pillére az alakzatfelismerés, a szegmentálás és a tárgykövetés az alábbi természetes kérdéseket próbálják megválaszolni: „*Mit látok a képen?*”, „*Mik tartoznak össze?*” és „*Hova megy a valami?*”.

Ezzel a képfeldolgozás témaköréből kiragadott laboratóriumi méréssel olyan gyakorlatot szeretnénk biztosítani hallgatóinknak, amin keresztül megízlelhetik és elsajátíthatják a bevezetőben felsorakoztatott példák alapjául szolgáló módszerek egy részét. Ehhez szükséges, hogy a hallgató rendelkezzen az alapvető optikai, geometriai és lineáris algebrai ismeretekkel.

1.1. A kamerakalibráció formalizmusa

Feladatul tűzzük ki, hogy leírjuk a minket körülvevő tér \mathcal{R} és az i -ik kamerából közvetített képi tér \mathcal{P}_i közötti kapcsolatot. Ezt a leképezést számos paraméter jellemzi. A paramétereket szokás belső és külső tényezőkre bontani úgy nevezett *intrinsic* és *extrinsic* paraméterekre. Az *intrinsic* paraméterek leírják a kamera belső jellemzőit, fókusz távolság, pixel-méret arány, torzítást jellemző paraméterek. Az *extrinsic* paraméterek például számot adnak a kamera tárgy terében értelmezett helyzetéről.

A leképezést meghatározó problémát, a könnyebb tárgyalás kedvéért és az elfogadható pontosságú eredmények miatt, első közelítésben lineárisnak tekintjük. Érdekes a problémát a projektív geometriában használatos homogén-koordinátákban leírni, mert így minden operátor mátrix alakban kezelhető. A homogén-koordinátákon értelmezett ekvivalencia reláció szerint két vektor ugyanazt a pontot írja le, ha egy nem nulla skalár szorzóval térnek el egymástól, azaz $(x, y, z, 1) \equiv (\alpha x, \alpha y, \alpha z, \alpha)$ ugyanazt a térbeli pontot jelöli.

Képzletben rögzítsünk egy koordináta rendszert a tárgy terében, \mathcal{R} -ben. Ezután válasszuk meg az i -ik kamera koordináta-rendszerét úgy, hogy a kamera optikai tengelye essen egybe a \mathcal{P}_i rendszer z -tengelyével, valamint a kamera fókusz pontja legyen az origóban, a részleteket lásd a 2(a). ábrán. Ezen koordináta rendszerek között a vektoralgebrában megtanult módon transzformációk segítségével váltunk. Homogén-koordinátákon értelmezhetjük az eltolás $T_{\mathbf{r}}$ és a forgatás O_{Ω} operátorait, ahol az $\mathbf{r} \in \mathbb{R}^3$ az eltolás vektora valamint $\Omega = (\omega, \varphi, \kappa) \in [-\pi; \pi]^3$ az Euler-szögek. A forgatásokat rendre az x, y majd a z tengely körül végezzük el.

$$T_{\mathbf{r}} = \begin{pmatrix} 1 & 0 & 0 & r_x \\ 0 & 1 & 0 & r_y \\ 0 & 0 & 1 & r_z \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad O_{\Omega} = \begin{pmatrix} o_{11} & o_{12} & o_{13} & 0 \\ o_{21} & o_{22} & o_{23} & 0 \\ o_{31} & o_{32} & o_{33} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

ahol a forgatás ortogonális operátorának elemei rendre:

$$\begin{aligned} o_{11} &= \cos \varphi \cos \kappa, \\ o_{21} &= \cos \varphi \sin \kappa, \\ o_{31} &= -\sin \varphi, \\ o_{12} &= \sin \omega \sin \varphi \cos \kappa - \cos \omega \sin \kappa, \\ o_{22} &= \sin \omega \sin \varphi \sin \kappa + \cos \omega \cos \kappa, \\ o_{32} &= \sin \omega \cos \varphi, \\ o_{13} &= \cos \omega \sin \varphi \cos \kappa + \sin \omega \sin \kappa, \\ o_{23} &= \cos \omega \sin \varphi \sin \kappa - \sin \omega \cos \kappa, \\ o_{33} &= \cos \omega \cos \varphi. \end{aligned}$$

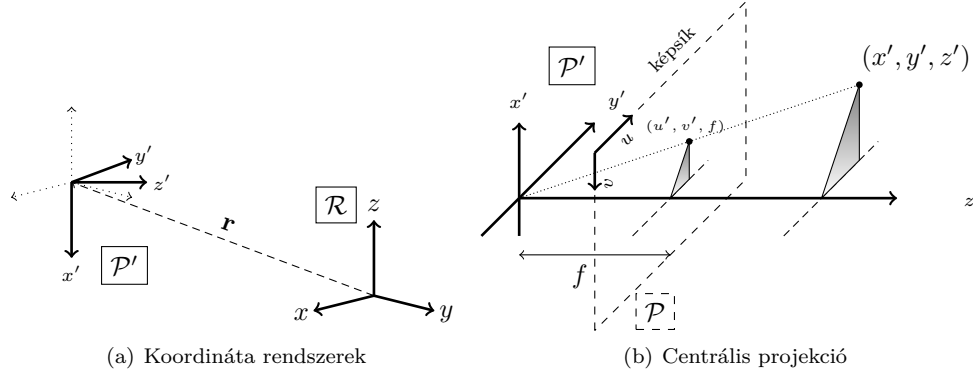
A probléma további tárgyalása az i -ik kamera rendszerében történik. A képalkotás folyamatát egyszerűsítjük, a lyukkamera modellt használjuk (*centrális projekció*), lásd a 2(b). ábrát. A modell értelmében a tárgy egy pontját egy egyenes segítségével össze kell kötni a kamera fókuszpontjával. Ez az egyenes dőfi a képsíkot (*fókuszsíkot*), amit a kamera rendszerében a $z = f$ egyenlet ír le. A képletben szereplő f a fókusz távolság a közös SI-beli távolság mértékegységekben. A hasonló háromszögeket jellemző arányossági szabályok értelmében a vetítés során a következő teljesül:

$$\Pi : (x, y, z, 1) \mapsto \frac{f}{z}(x, y, z),$$

aminek operátora a következő alakban írható:

$$\Pi_f = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

A tárgy vizsgált pontjának fókuszsíkon megjelenő vetületét ezután nyújtási transzformációval átskálázzuk, annak érdekében, hogy a távolság mértékegységeket



2. ábra. (a) A tárgytérhez (\mathcal{R}) és a kamerához (\mathcal{P}') rendelt koordinátarendszerek közötti kapcsolat. A kamera fókuszpontja \mathbf{r} helyzetvektorral jellemezhető \mathcal{R} -ben. A kamera optikai egyenese a \mathcal{P}' rendszer z tengelyével esik egybe. (b) A centrális projekció szemléletes jelentése. A tárgy egy kitüntetett pontja és a kamera fókuszpontja között definiált egyenes dőfi a kamera fókuszsíkját. A dőfés pont a tárgypont képe.

sége képpontokban legyenek értelmezve. Végezetül egy transláció segítségével a képek jobb felső sarkába mozgatjuk az origót.

$$S_{\mathbf{a}} = \begin{pmatrix} \alpha_x & \delta & 0 \\ 0 & \alpha_y & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \Lambda_{\rho} = \begin{pmatrix} 1 & 0 & \rho_u \\ 0 & 1 & \rho_v \\ 0 & 0 & 1 \end{pmatrix},$$

ahol $\mathbf{a} = (\alpha_x, \alpha_y, \delta)$ paraméterek a léptékváltásért felelősek. Ez modell leírás megengedi, hogy a kamerába épített CCD fényérzékelő lemez pontjai nem tökéletesen négyzet alakúak, az x illetve az y irányú léptékváltás paraméterei α_x és α_y , míg a δ változó a nyírást írja le, azaz amikor a képi rendszerben az x és y tengelyek nem merőlegesek egymásra. A lencse egyéb hibáitól eltekintünk. A teljes leképezés mátrixa ezek szerint:

$$\mathcal{R} \xrightarrow{O \circ T} \mathcal{P}' \xrightarrow{\Lambda \circ S \circ \Pi} \mathcal{P},$$

$$M = \Lambda_{\rho} S_{\mathbf{a}} \Pi_f O_{\Omega} T_{\mathbf{r}}. \quad (1)$$

M egy 3×4 mátrix, aminek elemei a fent bevezetett *intrinsic* és *extrinsic* paraméterek függvénye, minden i kamerára más és más.

1.2. M elemeinek gyakorlati meghatározása

az 1.1. alfejezetben bevezettük az M leképezést, ami egy $\mathbb{R}^4 \rightarrow \mathbb{R}^3$ homogén leképezés, az i -ik kamera sajátja. Feladat, hogy M elemeit gyakorlatilag meghatározzuk. Ehhez szükségünk van N darab ismert helyzetű \mathcal{R} -beli tárgypontra. Jelölje a tárgypontok halmazát $\{\mathbf{r}_1, \dots, \mathbf{r}_N\}$. A tárgypontok vetületei $\{\rho_1, \dots, \rho_N\}$.

Ezen vektorokkal jellemzett pontok homogén reprezentánsait jelöljük hullámos felülvonással ($\tilde{\cdot}$). Minden i pontra igaz az alábbi összefüggés:

$$\sigma_i \begin{pmatrix} u_i \\ v_i \\ 1 \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix} \cdot \begin{pmatrix} x_i \\ y_i \\ z_i \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \\ \mathbf{m}_3 \end{pmatrix} \cdot \begin{pmatrix} x_i \\ y_i \\ z_i \\ 1 \end{pmatrix},$$

ahol $\tilde{\mathbf{r}}_i = (x_i, y_i, z_i, 1)^\top$ és $\tilde{\rho}_i = (u_i, v_i, 1)^\top$ és \mathbf{m}_i az M operátor i -ik sora. Mivel homogén-koordinátákkal dolgozunk σ_i erejéig meghatározott az egyenletrendszerünk. Megtehetjük, hogy M operátort konstans skalárral bővítjük, például $m_{34} \neq 0$ elem reciprokával, mintegy beletranszformáljuk minden σ_i -be ezt az értéket. Látszik továbbá, hogy $\sigma_i = \mathbf{m}_3 \cdot \mathbf{r}_i$ skalárszorozattal egyezik meg, ahol már $m_{34} = 1$.

Ezen a ponton a következőt kell megjegyeznünk. M 11 eleme ismeretlen, és minden új referencia pont valójában két új egyenlettel bővíti az összefüggéseket. Ezek szerint legalább 6 referencia pont szükséges M elemeinek meghatározásához.

Az együttthatókat kereshetjük a legkisebb négyzetek módszerével. Defináljuk a χ^2 hiba függvényt az alábbi módon:

$$\chi^2 = \sum_{i=1}^N [(\mathbf{m}_1 \cdot \mathbf{r}_i - \mathbf{m}_3 \cdot \mathbf{r}_i u_i)^2 + (\mathbf{m}_2 \cdot \mathbf{r}_i - \mathbf{m}_3 \cdot \mathbf{r}_i v_i)^2].$$

A χ^2 szemléletesen azt méri, hogy a tárgy pont vetülete és a mért vetületpontok mennyire térnek el egymástól. Ebben a definícióban az eltérést az Euklidészi-távolság ragadja meg. M megoldását ott várjuk, ahol χ^2 -nek szélső értéke van, azaz minden ismeretlen m_{ij} -re a képzett parciálisan derivált kifejezése eltűnő. Az M elemeire így levezetett új egyenletrendszer megoldható Gauss-eliminációval.

1.3. M dekomponálása és a jellemző paraméterek becslése

Az 1. definícióból látszik, hogy a kamera belső paramétereit megadó $\Omega := \Lambda \Pi$ egy felső háromszög mátrix. A forgatás O operátora ortogonális, azaz $O^\top O = \mathbf{1}$. Jelölje $X := \Omega O$ -t.

$$M = X T_{\mathbf{r}} = (\mathbf{m}_1 \ \mathbf{m}_2 \ \mathbf{m}_3 \ \mathbf{m}_4).$$

T alakjából és ezen összefüggés szerint az X és az M operátor első 3 oszlopából alkotott 3×3 -as mátrix megegyezik. Vezessünk be jelölést erre a mátrixra $N := (\mathbf{m}_1 \ \mathbf{m}_2 \ \mathbf{m}_3)$. Vegyük észre továbbá, hogy Ω negyedik oszlopa zérus elemeket tartalmaz csak, így X mátrixa is ilyen. Ebből az következik, hogy $X \tilde{\mathbf{r}} = \mathbf{m}_4$. Szorítsuk meg az $\tilde{\mathbf{r}}$ homogén vektort a természetes \mathbb{R}^3 térre. Így lehetőségünk nyílik arra, hogy közvetlen becslést származtassunk a kamera tárgytérbeli pozíciójára \mathbf{r} -re:

$$N \mathbf{r} = \mathbf{m}_4 \quad \longrightarrow \quad \mathbf{r} = (\mathbf{m}_1 \ \mathbf{m}_2 \ \mathbf{m}_3)^{-1} \mathbf{m}_4.$$

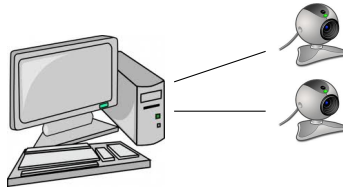
A fenti eljárással M operátorból az eltolást (T -t) sikerült kitranszformálni. Következő lépés a forgatás meghatározása. Ezt az X operátor RQ-felbontásával tesszük meg, ami pont az ΩO szorzatra bont, így gyakorlatilag leválasztottuk az összes *extrinsic* paramétert, Ω már csak a kamera belső tulajdonságaitól ($f, \alpha_x, \alpha_y, \delta, \rho_u, \rho_v$) függ.

1.4. Egy térbeli pont előállítás a lenyomatok ismeretében

Ebben a fejezetben körvonalazzuk, miként lehet a tárgy egy pontjának térbeli helyzetéről becslést adni, több egy időben exponált kamerakép ismeretében. A probléma formálisan a következő, szeretnénk $\mathbf{r} \in \mathcal{R}$ adott tárgypont helyzetét leíró vektort meghatározni. Ennek a pontnak a képeken meg kell keresni a lenyomatait, az i -ik képen jelölje ezt ρ_i . Az 1.1. fejezet alapján felbontottuk M leképezést, azaz az i -dik kamerát $M_i = \Omega_i O_i T_i$ jellemzi. Természetesen M_i nem invertálható, de az Ω_i -beli nyújtások és eltolás igen. Azt alkalmazva ρ_i -re és kiegészítve, hogy a w komponens az f távolságban legyen kapjuk a vetítő vonal dőfspontját $\tilde{\rho}_i$ -t. A vetítő egyenest két pontja jellemzi, így a kamera fókuszpontja, ami egybeesik a kamera koordinátarendszerének origójával valamint ez a dőfspont ($\tilde{\rho}_i$). Ezeknek ösképet megkaphatjuk a $T_i^{-1} O_i^T$ operátort alkalmazva. Azaz leírjuk a tárgy terében a vetítő egyenest. Minden i vetületre előállítva a vetítő egyeneseket, a keresett tárgypont ezeknek a metszete. A hibák és a mérést jellemző zaj miatt ezek az egyenesek nem minden esetben metszik tökéletesen egymást. Tehát célszerű azt a pontot választani, ami valamilyen értelemben legközelebb esik az összes vetítő egyeneshez. Például tekintsük a tárgypontot olyannak, ami a vetítő egyenesektől mért távolságok összegét minimalizálja.

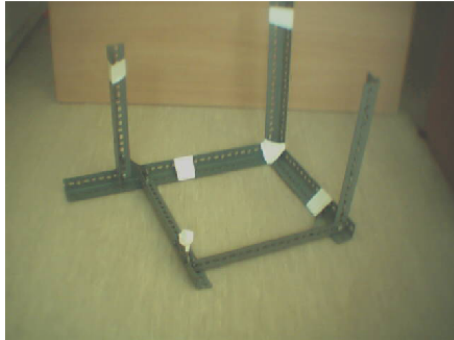
2. Mérési összeállítás

A labormérés során a képi információkat a személyi számítógép USB csatolójára csatlakoztatott kamerák segítségével készítheti el, majd a képi adatok transzformációját, feldolgozását is a számítógépen végzi. A mérés sematikus rajza a 3. ábrán látható.

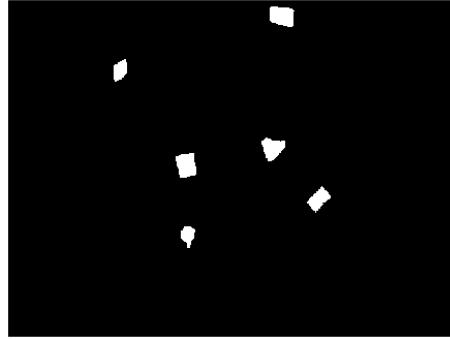


3. ábra. A mérési összeállítás. A mérés lelke a személyi számítógép, melyhez kettő hagyományos kamerát kapcsolunk az USB csatolón keresztül.

Az első legfontosabb feladat a 3. ábrán felvázolt rendszer kalibrációja. A kalibráció során információt kapunk a képalkotó rendszert jellemző belső paraméte-



(a) Referencia rendszer



(b) Objektumok

4. ábra. A valódi térünk és a képi koordináták között meg kell keresni a kapcsolatot. Az ezzel kapcsolatos paramétereket a kamerák kalibrációjához használt referencia objektum segítségével határozzuk meg.

rekről (a kamera fókusztávolsága, torzítása, ...) és a kamera térbeli elhelyezkedéséről (koordináta-rendszerek eltolási és forgatási transzformációi) valamint a valós tér és a képi térben használt természetes mértékegységek közötti kapcsolatról.

Az 1.1. részben ismertetett kamerakalibrációhoz referencia pontokat kell felvennünk a fizikai térünkben. A 4(a). ábrán a referencia pontjainkat egy merev keretre erősítettük. A pontok felfedezését megkönnyítendő a referencia pontokat világos színűnek választottuk, amit könnyű keresni, sötét háttér mellett. A színes kép szürke skálás konverziója és a megfelelő küszöbök beállítása után előállítható a 4(b). ábrán látható bit-térkép, melyen csak a referencia pontok helyén vesznek fel magas (1) értéket, a háttér alacsony (0). Feladat a csoportok elkülönítése és mindegyik csoporthoz hozzárendelni egy-egy képi koordinátát.

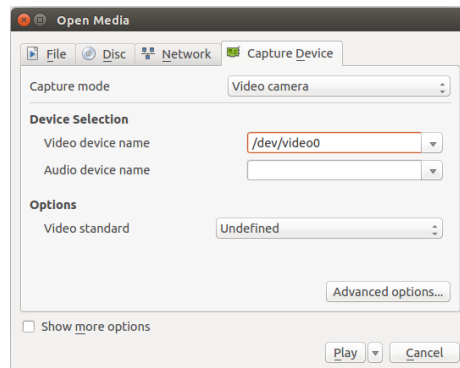
2.1. Szoftver komponensek

A kitűzött problémák bemeneteit képkockák alkotják. A képkockákban tárolt adatok feldolgozása és kiértékelése a számítógép segítségével történik `python` nyelven. Ezeknek a képkockáknak a kamerából történő kiolvasására nagyon sokféle programot használhatunk, mi magunk is készíthetünk eljárásokat hozzá. A gépi képfeldolgozást segítő, nagyon hasznos eljárásokat tartalmaz az `opencv` könyvtár.

cite

Sok esetben segíti a megfelelő képkocka kiválasztását az élőképes előnézet. Élőnézetben a tárgyra tudjuk irányítani a kamerát, és lehetőség nyílik a kamerára szerelt lencsék segítségével a tárgy képét élesre állítani.

A kamerák vezérléséhez, az élőkép megjelenítésére és a használni kívánt képkockák rögzítésére használjuk a `vlc` programot. A menürendszer segítségével válasszuk ki a használni kívánt forrást. A `Media > Open capture device...`



5. ábra. A vlc programban válasszuk ki azt a kamerát, aminek a képeivel dolgozni szeretnénk.

menüelemmel az 5. ábrán látható panel ugrik fel, és a **Video device name** legördülő menüjéből válasszuk ki a kamerát. A **Play** gomb nyomására a kamera inicializálódik, és folyamatos élőképet kapunk.

A **Video > Take snapshot** menüponttal a `~/Pictures/` könyvtárban rögzítheti a kameraképet. Ne felejtse el a jegyzőkönyv könyvtárába másolni a szükséges fájlokat.

Feladatok

A feladatok elvégzését igazolandó jegyzőkönyvében tisztán naplózza munkáját, beleértve eljárásainak elvét, implementációit valamint eredményeit. Ne felejtse el jegyzőkönyvéhez csatolni, azaz a jegyzőkönyvet tartalmazó munkakönyvtárba másolni a feladatok kidolgozására elkészített bemeneti képállományokat is.

Ismerkedjen meg a mérési összeállítással.

Kamerakalibráció

A perspektíva vizsgálata

A mérés helyszínén adott egy pepita rács. A kísérlet során a kamerával készítsen erről a rácsról felvételeket úgy, hogy a kamerától különböző távolságokra helyezi le azt. Ügyeljen arra, hogy lehetőleg mindig a legnagyobb felületet lássa a kamera, azaz essen a kamera optikai tengelye és a rács közepén átmenő normális (a lehetőségek szerint) egybe. A képmintát elemezve minden felvételen határozza meg a rács karakterisztikus távolságát. Mit vár és mit tapasztal, ha a rácsot jellemző karakterisztikus távolságokat megvizsgálja a rács-kamera távolság függvényében. Tapasztalatairól készítsen ábrát is!

A kamera leképezési mátrixának meghatározása

A munkaterületen Salgo elemektől koordinátarendszert építettünk. Ezen markerek találhatóak, melyekről készítsen felvételt a kamera segítségével. Mérje meg mérőszalag segítségével a markerek térbeli helyét és foglalja táblázatba. Határozza meg ugyanezeknek a markereknek a helyét a mintaképen és vezesse azok képkordinátáit is fel a táblázat megfelelő helyére. A jegyzetben tárgyalt egyenletek segítségével határozza meg a kamera leképezési mátrixát. Ellenőrzésképpen ábrázolja a 3D térben feszülő koordinátatengelyeket a képen.