

Testkulcspontok becslése FMCW radarképek segítségével

vajdad4m

Vajda Ádám

Felkészítő tanár: Dr. Vajda Róbert

ELTE Radnóti Miklós Gimnázium, 1146 Budapest, Cházár András u. 10.

1. Bevezetés

1.1. Célkitűzés

Kutatási célom egy olyan mesterséges intelligencia modell megtervezése és betanítása volt, amely képes valós időben, vizuális okklúzióban felismerni egy személy testének kulcspontjait radarképek alapján.

1.2. Téma aktualitása

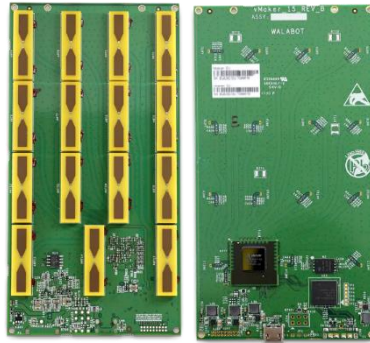
A mesterséges intelligencia taktikai és aszimmetrikus előnyt biztosíthat a közelharcban. Az emberi test kulcspontjainak (a továbbiakban: póz) becslése a számítógépes látás egyik sokat kutatott területe. Többek között a megfigyelésben, a gesztusfelismerésben és a virtuális valósággal való interakciókban használható. Hasznos információkat szolgáltatathat többek között a természeti katasztrófák utáni mentési műveletek során is.

Számos olyan helyzet van, amikor a hagyományos, kamera alapú pózfelismerés nem alkalmazható a látási viszonyok, vagy esetleg a kilátást akadályozó akadály miatt. A radaralapú pózfelismerés viszont megoldja ezt a problémát. Ugyanakkor szükségünk van egy olyan módszerre, amellyel a póz adatokat kinyerhetjük a radarképekből.

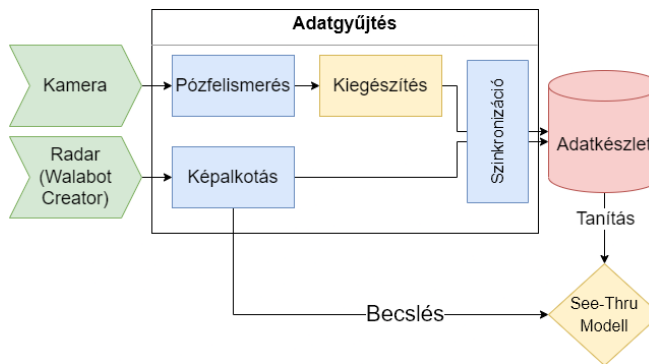
1.3. Megvalósítási terv

Az FMCW radarok hatékonynak bizonyulnak a kis hatótávolságú képalkotás megoldásában. Ebből a célból a Walabot „Creator” radarját választottam.

A kutatás megvalósításához saját radar- és póz adatokat kellett gyűjtenem, amihez egy radar-kamera szinkronizációs rendszert fejlesztettem. Kerestem egy ígéretes módszert a póz kulcspontjainak a kameraképekből való kinyerésére. Ki kellett dolgoznom egy modellt, hogy kiegészítsem a hiányzó kulcspontokat. Mindezek segítségével pedig egy olyan modellt terveztem és tanítottam fel, amely a radarképek alapján kiszámítja a kulcspontokat.



1. ábra: Walabot „Creator” Radar



2. ábra: Megvalósítási terv

2. Probléma megoldásának menete

2.1. Radar

A radar paramétereit a kamera paramétereinek (90°-os látószög, 640×480-as felbontás) figyelembevételével optimalizáltam. Látóterének 60°-ot választottam, hogy megakadályozzam a határon túli méréseket. Ismert, hogy a pózfelismerő módszer elveszti pontosságát, ha a célpont a képkocka szélén van, ezért a pózfelismerést, így a radart a képkocka belső részére kell fókuszálni. Ez a 15°-os különbség mindkét irányban, úgy tűnik, hogy megoldotta ezt a problémát.

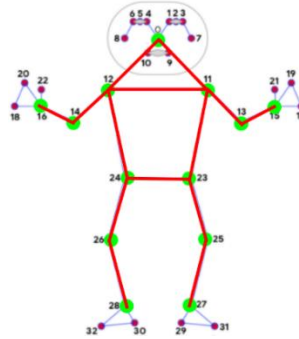
2.1. Kamera és pózfelismerés

A testkulcspontok felismeréséhez a Google BlazePose (Bazarevsky, et al., 2020) módszerét használtam, amely 640×480-as méretű képekből 32 egyedi kulcspontot képes felismerni és követni. Ebből a 32 kulcspontból 13

	R	θ	ϕ
Res / Felbontás	5 cm	1°	1°
Min	100 cm	-30°	-30°
Max	250 cm	30°	30°

2. táblázat: A radar mérési paraméterei

főkulcsponthoz választottam ki követésre és mentésre, a többi 19-et elvetettem. Ez azért szükséges, mert a radar felbontása nagyobb, mint a törölt pontok közötti távolság (például ujjak közötti távolság), ezért teljesen felesleges számítási kapacitást igényelnének.



4. ábra: Kiválasztott kulcsponthoz

2.2. Pózkiegészítés

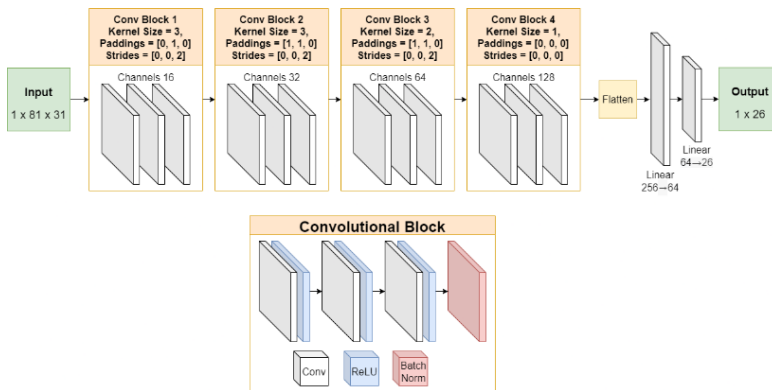
El kell ismerni, a fent említett módszerrel történő rögzítéskor nem minden kulcsponthoz megállapítható. Ennek a problémának a megoldására egy olyan modellt használtam, amit teljesen rögzített kulcsponthoz képeztem ki. A modell veszteségfüggvénye az alábbi:

$$L = \sum_{i=1}^n \sqrt{[640(R_{x_i} - P_{x_i})]^2 + [480(R_{y_i} - P_{y_i})]^2}$$

A veszteségfüggvény az euklideszi távolságok pixeltéri összege a megjósolt és a címkézett kulcsponthoz mérve. L a kapott veszteség értéke, R_x , illetve R_y a valós póz x és y koordinátáinak vektora, P_x és P_y , pedig a becsült póz koordinátáinak vektora. A pózkiegészítő modell architektúrája egy 0,2-es meredekségű Leaky ReLU aktivációs funkciós MLP modell.

2.3. Pózkinyerés radarképekből

Az adatgyűjtést elvégezve ~1700 póz-radarkép pár állt rendelkezésemre, ezek különböző helyzeteket, illetve különböző cselekvéseket (sétálás, ülés, állás) rögzítő mérések. A radarképekből a test kulcspontjait kinyerő modell egy konvolúciós neurális hálózat, amelynek két lineáris réteg a feje. A CNN-ek használatának számos előnye van. Először is, jól működnek a képekkel kapcsolatos feladatokban, másodsor pedig, kevés paramétert igényelnek. A modell összesen 1,7 millió paramétert tartalmaz, ami gyors következtetési időt eredményez. Az architektúrája egy VGG-16 (Simonyan & Zisserman, 2014)-hoz hasonló modell, módosított kimeneti fejjel.



5. ábra: A modell architektúrája

3. Elért eredmények

Az eredmények azt mutatják, hogy kulcspontonként átlagosan **7,5 cm**-en belüli pontosságot érhetünk el. Megfigyelhető, hogy ezt a feladatot egy embernek szinte lehetetlen lenne elvégezni, de egy mesterséges intelligencia modell képes nagy pontossággal végrehajtani. A radarképek világosan mutatják a koncentrációt, ahol az ember áll. Ez alapján a személy radartól való távolsága is mérhető.

4. Összefoglalás

Munkám során bemutattam egy olyan rendszert, amely képes az emberi test tájékozódási pontjainak felismerésére radarképekből. Ennek megvalósításához két mély tanulási modellt hoztam létre. Az egyik képes nagy pontossággal megjósolni a hiányzó emberi kulcspontokat egy adott pózban. A másik modell 13 különböző kulcspontot tud reprezentálni a radarképekről.

A bemutatott rendszer alkalmazható terrorelhárításnál, természeti katasztrófák utáni mentési műveleteknél, repülőgépek biztonsági szintjének növelésére, gesztusvezérlésben és egyéb biztonsági rendszerek fejlesztésében.