



M Ű E G Y E T E M 1 7 8 2

Budapesti Műszaki és Gazdaságtudományi Egyetem

Villamosmérnöki és Informatika Kar

Irányítástechnika és Informatika Tanszék

3D alakfelismerő módszerek tapintható felhasználói felületekhez

PHD ÉRTEKEZÉS

SZEMENYEI MÁRTON

Témavezető

Dr. Vajda Ferenc

2021. február 13.

Tartalomjegyzék

1. Bevezető	2
2. Célok és módszertan	3
2.1. Célkitűzés	4
2.2. Módszertan	4
3. Új tudományos eredmények	6
3.1. Forma osztályozás	7
3.2. Dimenzió redukció	9
3.3. Jelenetek berendezése	12
4. Eredmények összegzése	15

1 Bevezető

A felhasználói felületek a számítógépekkel és más eszközökkel való interakciónk elsődleges módja, így ezek hatékonysága és használhatósága az informatika területének központi kérdése. Megfigyelhető ugyanis, hogy a három nagy informatikai forradalom közül kettő egy új interakciós technika megjelenéséhez köthető: a grafikus felhasználói felületek és az egér elterjedése jelentős szerepet játszott a személyi számítógépes elterjedésében a 80-as években, valamint a jó minőségű érintőképernyők megjelenése fontos tényező volt az okostelefonok és tabletek népszerűségében.

Jelenleg az egér és a billentyűzet a személyi számítógépekkel történő interakció elsődleges felülete, míg a mobil eszközök elsősorban érintőképernyős technológiát alkalmaznak. Belátható, hogy mind az egér, mind az érintés-alapú felületek népszerűségének egyik legfőbb oka, hogy használatukhoz az ember számára természetes metaforák megértése szükséges, mivel a téridő és objektumok manipulációjának ösztönös megértésére támaszkodnak. A természetes metaforák használatának következményeképp könnyen tanulható felületet kapunk, amely - amíg megbízhatóan működik - a felhasználók számára rendkívül hasznos eszköz.

Észrevehető azonban, hogy mindkét tárgyalt interakciós technika alapvetően kétdimenziós, ami megnehezíti a magasabb dimenziós manipulációt igénylő környezetek használatát. Ez egy jelentős hátrány, mivel a valós életben gyakran találkozhatunk három-, illetve négydimenziós struktúrákkal. Ez a probléma indokolja további újszerű, sokdimenziós felhasználói felületek kutatását és fejlesztését.

A tapintható felhasználói felületek a tudomány egyik intenzíven kutatott területe volt az utóbbi évtizedben. E rendszerek célja, hogy lehetővé tegyék a felhasználó számára, hogy virtuális objektumokat valós, fizikai tárgyakon keresztül manipuláljon, ez által egy természetes felületet szolgáltatva. A tapintható felületeken belül a tapintható kiterjesztett valóság egy külön kiemelendő terület, melynek lényege a két terület kombinálása a kiterjesztett valóság felületbe becsatolt valós objektumok segítségével.

2 Célok és módszertan

Különböző formájú objektumok más manipulációs technikákat igényelnek, és eltérő érzetet adnak érintéskor. A látás és a tapintás érzékei közti inkonzisztencia neurális konfliktushoz vezethet, amely nagymértékben csökkenti a felhasználói élményt. Azonban, ha a virtuális és valós objektumokat alakbeli hasonlóság alapján párosítjuk, akkor ez a konfliktus csökkenthető. Az így kapott párosító eljárás segítségével létrehozható egy Adaptív Kevert Valóság (AKV) rendszer: Egy olyan virtuális környezet, amely képes a virtuális objektumok intelligens elrendezésével az adott valós környezethez alkalmazkodni.

Jelen értekezés témája egy olyan újszerű algoritmus, amely képes a virtuális és valós objektumok párosítását meghatározni. A párosítás fő kritériuma, hogy a valós és virtuális objektumok tulajdonságai hasonlóak legyenek, hogy a felhasználó pont úgy manipulálhassa a valós objektumot, mint ahogy azt a virtuálissal tenné. Természetesen sok fizikai tulajdonsága (pl.: tömeg, felület durvasága) nem becsülhető vizuálisan, így a megalkotott módszer csupán a formát és a méretet használja.

Megjegyzendő, hogy a kevert valóság rendszereknek további követelményei is vannak: Például bizonyos objektumok feltétlenül szükségesek, mely esetben a rossz minőségű párosításokat is el kell fogadni. Továbbá, bizonyos esetekben előnyös lehet egy virtuális objektumból minél több példányt behelyezni, míg máskor egy is elég. Végezetül lehetnek virtuális objektumok, amelyeket érdemes egymáshoz közel helyezni, mely esetben az algoritmust az előnyös elrendezésért jutalmazni kell.

Ily módon az AKV rendszerek képesek csökkenteni a felhasználó által végzendő előkészületi munkákat: Az AKV rendszer *fejlesztője* elkészíti a virtuális objektum kategóriákat, betanítja a rendszert, hogy az ezekhez tartozó valós objektumokat jól felismerje, és beállítsa az egyes virtuális kategóriák tulajdonságait és szükségességét. Ezeket felhasználva a rendszer megkéri a felhasználót, hogy járja körbe a valós jelenetet a 3D rekonstrukció elvégzése céljából, majd automatikusan meghatározza a virtuális objektumok helyét.

2.1. Célkitűzés

Nyilvánvaló, hogy a javasolt párosító eljárás alabeli hasonlóságra kell, hogy alapuljon. Bár számos különböző módszer létezik alak alapú objektumpárosításra, a jelen munka célja egy olyan módszer megalkotása, amely nem igényel a priori ismereteket a virtuális objektumokról, hanem képes ezeket megtanulni egy címkézett objektum adatbázis segítségével. Továbbá, hogy ne kelljen az kisméretű (lokális) és a globális tulajdonságok fontosságáról feltételezéseket tenni, olyan leíró módszert választunk, amely mindkettőt magába foglalja, így maga a döntés a tanuló módszerre lesz bízva. Mivel egy jelenet formája strukturált objektumként írható le a legjobban, az értekezés fő fókuszja az ilyen komplex formákat leíró strukturált objektumokon történő tanuló algoritmusok megalkotása lesz.

Fontos azonban megjegyezni, hogy az alakfelismerés önmagában nem nyújt teljes megoldást a problémára, mivel a fejlesztő által támasztott további követelmények miatt az optimális címkék egymástól is függenek. A probléma megoldására valamilyen globális optimalizálási sémát kell bevezetni, amely az összes objektumhozrendelést közösen optimalizálva eljut egy optimális kompromisszum elrendezéséhez.

2.2. Módszertan

Az értekezésben bemutatott módszerek javarészt a gépi tanulásra és más heurisztikákra támaszkodva próbálnak nehezen definiálható kritériumokat kielégíteni (mint például a formabeli hasonlóság). Ráadásul olyan problémára alkalmazzuk őket, amelynek nincs garantáltan létező megoldása (elképzелhető például, hogy kevesebb valós objektum van, mint szükséges virtuális). Következésképp, az algoritmus helyes működése minden lehetséges bemenetre nem bizonyítható hagyományos módszerekkel.

Ezen okokból kifolyólag a javasolt módszerek hatékonyságát empirikus módon, számos különböző adatbázis segítségével igazoljuk. Ez azonban felvet egy újabb problémát: mivel a probléma meglehetősen ritka (legjobb tudomásunk szerint a javasolt alkalmazás új), így nem állnak rendelkezésre nyilvános adatbázisok, amikkel a tesztelés elvégezhető.

A probléma megoldására saját adatbázisokat készítettem, amelyekben irodai asztalokon gyakran található valódi objektumtípusok szerepelnek. Az egyik adatbázisban Blender-ben készített szintetikus képek, a másokban pedig valós képekből készült 3D

rekonstrukciók találhatóak. Nyilvánvaló azonban, hogy két adatbázis nem elegendő, ezért ezeken felül számos további szintetikus adatbázist készítettem. Ezek az adatbázisok automatikusan lettek generálva, különböző véletlenszerűen választott hiperparaméterekkel. Belátható, hogy több, eltérő struktúrájú adatbázis jobban bemutatja a megalkotott módszer univerzalitását és robusztusságát, mint egy, sok adatot tartalmazó adatbázis.

Amint a módszereket az összes adatbázison kiértékeljük, statisztikai tesztek végzünk a saját és a korábbi state-of-the-art (SOTA) módszerek eredményei között. Mivel a módszerek ugyanazokon az adathalmazokon lettek kiértékelve, a párosított mintás t-próbát alkalmazzuk, hogy kiértékeljük a módszerek hatékonyságai közti különbséget. Természetesen több adatbázis esetén a statisztikai teszt eredménye is robusztusabb lesz.

Létfontosságú azonban megjegyezni, hogy a statisztikai próbák hagyományos, frekventista verziói (pl.: a Student-féle t-próba) csak az adat valószínűségét határozzák meg a nullhipotézis feltételezése mellett. Súlyos (és túl gyakori) tévedés ebből a null (vagy bármely más) hipotézis valószínűségére bármilyen következtetést levonni. Sajnálatos módon, számunkra pontosan egy bizonyos hipotézis valószínűsége vizsgálendő.

Ezen okokból kifolyólag a páros mintás t-próba Bayesi változatát alkalmazzuk, amely lehetővé teszi, hogy a hatás (vagyis a javasolt módszer által okozott javulás) valószínűségére, valamint annak a 95%-os konfidencia intervallumára következtessünk. A Bayesi változat használatának ára, hogy a priori eloszlást kell állítanunk a hatásra, amelyre egy széles normális eloszlást használunk, mivel ez egy nulla közepű, szimmetrikus függvény, ami nem vezet be bias-t a tesztbe.

3 Új tudományos eredmények

Az értekezésben bemutatott algoritmus első lépése a szegmentáció és alakleírás. Objektumdetektáló algoritmusok, amelyek képesek több objektum egyidejű észlelésére általában használnak valamilyen szegmentáló eljárást egymástól elkülönülő objektumjelöltek előállítására, melyeket egy ezt követően osztályozzák. Ez a módszer jól használható olyan esetekben, ahol az objektumok különválasztása egyszerű.

Beltéri jelenetekben azonban problémás lehet a szegmentáció, ugyanis ilyen esetekben gyakoribb a sűrű elrendezés. Éppen ezért a javasolt módszer egy más megközelítést alkalmaz: a bemeneti 3D jelenetet először építőelemekként értelmezhető primitív formákra bontjuk. ezt követően ezeket a primitív formákat egyesével osztályozzuk, az objektumokat pedig a közeli megegyező címkéjű szegmensekből építjük.

Mivel az objektumok közti geometria kapcsolatokat érdemes figyelembe venni, ezért a primitív formákból egy gráfot konstruálunk, majd minden csomópontot egy vektortérbe ágyazzuk be, így minden csomóponthoz egy olyan jellemző vektort rendelünk, amely a csomópontot és annak lokális kontextusát is leírja. A kezdeti osztályozást az így készített leírók segítségével végezzük.

Az eljárás következő lépésében egy diszkriminancia analízis algoritmus segítségével csökkentjük a leíró vektor dimenziószámát, így redukálva a tanuló osztályozó algoritmus komplexitását és erőforrásigényét. A dimenzióredukációs eljárás célja, hogy a helyes osztályozáshoz szükséges jellemzőkön felül az *egyetlen objektumon belüli* különböző csomópontok elválasztásához használható jellemzőket is megtartsuk. Ezen csomópontok felismerése ugyanis felhasználható a valós tárgy pozíciójának és orientációjának becslésére.

Az módszer végső lépésében felhasználjuk a kezdeti osztályozás eredményét egy globálisan optimális elrendezés meghatározására, amely minden egyéb követelményt is képes figyelembe venni. Ehhez a lépéshez egy genetikus algoritmust használunk, újszerű, a problémára kifejlesztett genetikus operátorokkal. Ebben a lépésben a kontextusbeli követelmények is kielégíthetők, lehetőséget nyújtva a fejlesztőknek az egyes objektumok közelségének előírására.

A célom az értekezés során, hogy megmutassam, hogy a megalkotott módszerek használhatók alakfelismerésre és az elrendezés meghatározására, valamint, hogy a korábbi módszerekhez lépest teljesítménybeli javulást is eredményeznek.

3.1. Forma osztályozás

Ahogy korábban említettük, az objektumpárosító eljárás első lépése egy egyszerű alakfelismerés a 3D jelenet részein, amely egy előzetes, független párosítást nyújt, amit későbbi lépésekben finomítunk tovább. A jelenlegi megközelítés ezt egy osztályozási problémaként kezeli: Az egész jelenetet részekre bontjuk, a részeket pedig aszerint osztályozzuk, hogy azok melyik virtuális objektumhoz tartoznak. Az osztályozási jóságokat a későbbi globális optimalizációs lépés fogja felhasználni inicializációhoz.

Az osztályozási pontosság növelésére egy újszerű gráf csomópont beágyazási keretrendszert alkottam. Ezek a módszerek tetszőleges gráfokat képesek kezelni struktúrabeli megkötések nélkül, miközben alkalmazhatók olyan gráfokra is, ahol a csomóponti és él súlyok vektorok. A megalkotott eljárások tovább javítják az osztályozási pontosságot a csomópontok lokális környezetének leírásával.

Az első megalkotott módszer egy explicit csomópont beágyazási keretrendszer, amely a gráfok spektrális felbontásán alapul. Mivel a spektrális felbontás csak részlegesen invariáns a csomópontok sorba rendezésére, ezért egy kezdeti egyértelmű sorba rendezés segítségével elérjük, hogy a leírók minden csomóponthoz egyediek legyenek. Ezt követően egy egyedi \mathbf{F} jellemző mátrixot konstruálunk minden csomóponthoz, amely a szomszédos csomópontok jellemzőit is tartalmazza az alábbi képlet alapján:

$$\mathbf{F} = \begin{bmatrix} T_{1,1} & T_{1,2} & \cdots & T_{1,N} \\ T_{2,1} & T_{2,2} & \cdots & T_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ T_{N,1} & T_{N,2} & \cdots & T_{N,N} \end{bmatrix} \quad (3.1)$$

$$T_{ij} = T(n_i, n_j, e_{1i}, e_{1j}, e_{ij}), \quad (3.2)$$

ahol T egy jellemző transzformációs függvény, n_i a i -edik csomópont, míg e_{ij} az i -edik és j -edik csomópontok közti él. N a maximális figyelembe vett lokális csomópontok száma. A jellemző transzformációs függvény tetszés szerint választható: az értekezésben tárgyalt problémához egy távolsággal súlyozott kernel függvényt javasolunk a csomóponti és él vektorok között. Az így kapott mátrix spektrális felbontásából már adódik a csomópont leírója.

A második javasolt módszer a véletlen-séte gráf kernel módosított változata. Mivel jelenleg csomópontokat és nem egész gráfokat kívánunk összehasonlítani, így a kiinduló valószínűség vektort úgy módosítjuk, hogy a séták mindig az összehasonlítandó

csomópontokból induljanak. A bejárt környezet méretét limitáljuk a séták maximális hosszának meghatározásával. A két gráf direkt szorzatát alkalmasint választott csomópont és él kernelek segítségével számoljuk.

Mindkét eljárást forma-gráf adatbázisok segítségével teszteltük, az osztályozási pontosságokat a beágyazás használata nélküli eredményekhez hasonlítottuk. Bayes-i t-próbák segítségével megmutattuk, hogy a megalkotott eljárások *szignifikánsan* javítják a kereszt-validációs pontosságot, továbbá a véletlen séta alapú módszer szintén javulást jelent az explicit beágyazáshoz képest. Másfelől az explicit módszer lényegesen gyorsabb, és *szignifikánsan* kevésbé hajlamos túltanulásra. Végezetül az explicit módszert neurális háló alapú baseline módszerrel is összehasonlítottuk, a teljes jelenetekre történő általánosítást ellenőrizve. Az eredmények azt mutatják, hogy míg az explicit beágyazás képes jól általánosítani, a neurális módszer csak a véletlen predikció esetében várható pontosságot tudta elérni.

Ezekén felül a megalkotott keretrendszerek elég általánosak ahhoz, hogy forma-gráfokon kívül más alkalmazásokban is használhatók legyenek. Jól megválasztott kernelek segítségével a módszerek alkalmazhatók 2D látás esetében is, például vizuális szóhalmaz típusú osztályozási, vagy deformálható rész-alapú detektálási feladatokra. Ezekben az esetekben a szomszédos vizuális elemek és a geometria figyelembe vétele a beágyazás során javíthatja az osztályozás pontosságát.

Végezetül, a módszerek nem limitáltak vektor csomóponti és él súlyokkal rendelkező gráfok esetére. Amíg megfelelő kernel függvények definiálhatók a csomópontok és élek között, addig a módszerek alkalmazhatók maradnak.

1. Tézis

Újszerű módszereket alkottam vektor-gráf csomópontok osztályozásának segítésére, amelyek az adott csomópont jellemzőin felül annak lokális kontextusát (más közeli csomópontokat és a köztük lévő geometriai kapcsolatokat) is felhasználják. Statisztikai tesztekkel igazoltam, hogy a módszerek javítják 3D forma-gráfok osztályozásának pontosságát. Ez az eredmény az alábbi részekből áll:

- (a) Új módszert alkottam vektor-gráf csomópontok vektortérbe történő beágyazására, amely a beágyazandó csomópont lokális kontextusát is elkódolja. Statisztikai tesztekkel igazoltam, hogy a módszer képes robusztusan általánosítani változó kontextusra, szemben korábbi módszerekkel.*
- (b) Kiterjesztettem a véletlen séta gráf kernelt vektor-gráf csomópontokon értelmezett kernellé.*

[1, 2, 3, 4, 5]

3.2. Dimenzió redukció

Az előző fejezetben bemutatásra került egy új eljárás vektor súlyú gráfcsozópon-
tok beágyazására és osztályozására, melynek használhatósága is bemutatásra került.
A javasolt módszer azonban nagy dimenziószámú leírókat eredményezhet, aminek
nagy számításigényű és túltanulásra hajlamos osztályozáshoz vezethet. Ebből ki-
folyólag célszerű lehet dimenzióredukciós módszereket alkalmazni a reprezentáció
tömörítésére.

Ez az alfejezet egy olyan módszert ismertet, amely segítségével strukturált kompo-
zit osztályokon végezhetünk diszkriminancia analízist. Ilyen osztályokban egyetlen
objektumhoz nem egy vektor, hanem vektorok rendezetlen halmaza, vagy gráfja
tartozik. Az egyes vektorokat az objektum *csomópontjainak* nevezzük. A mód-
szer feltételezi, hogy az egyes csomópontok normális eloszlásúak, azonban az egy
objektumon belüli csomópontok különböző eloszlásokból származnak.

Továbbá, az összes osztály összes objektumának csomópontjai ugyanabban a vektor-
térben találhatóak. Végezetül, a módszer nem szab korlátozásokat az egy objektumon
belüli csomópontok számának, így az osztályok és objektumok között is változhat.
Azonban nem áll rendelkezésre semmilyen címkézés a csomópontok objektumok közti
megfeleltetésére, információnk csupán arról van, hogy a csomópontok melyik osztályhoz és objektumhoz tartoznak.

A strukturált kompozit diszkriminancia analízis további követelménye, hogy az egy
objektumon belüli csomópontokat is el tudjuk választani, mivel ez orientáció bec-
sülésére felhasználható. Ezt a követelményt aránylag könnyen be tudjuk illeszteni a
létező módszerekbe, úgy, hogy egy második diszkriminációs mátrixot adunk hozzá
a kritériumhoz. Ez az objektumon belüli szóródási mátrix, amelyet a következő
képlettel számolhatunk:

$$\mathbf{S}_{wi} = \sum_{i=1}^C \sum_{j=1}^{N_i} \sum_{k=1}^{n_{i,j}} (\boldsymbol{\mu}_{i,j} - \mathbf{x}_{i,j,k})(\boldsymbol{\mu}_{i,j} - \mathbf{x}_{i,j,k})^T, \quad (3.3)$$

ahol C az osztályok száma, N_i az i -edik osztályban lévő objektumok száma, $n_{i,j}$ a
 j -edik objektum csomópontjainak száma, $\mathbf{x}_{i,j,k}$ pedig a k -edik csomópont az i -edik
osztály j -edik objektumában, $\boldsymbol{\mu}_{i,j}^{inst}$ pedig ennek az objektumnak a csomópontjainak
az átlaga. Hasonlóképp, a hagyományos LDA osztályok közötti szóródási mátrixa a
következőkép definiálható:

$$\mathbf{S}_{bcn} = \sum_{i=1}^C (\boldsymbol{\mu} - \boldsymbol{\mu}_i)(\boldsymbol{\mu} - \boldsymbol{\mu}_i)^T, \quad (3.4)$$

ahol $\boldsymbol{\mu}$ az összes csomópont átlaga, $\boldsymbol{\mu}_i$ pedig az i -edik osztály csomópontjainak átlaga. Ezt követően a strukturált kompozit diszkriminancia analízis (SCDA) kritériuma a következőképp írható fel:

$$\max_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{S}_{bci} \mathbf{w}}{\mathbf{w}^T \mathbf{S}_t \mathbf{w}}, \quad (3.5)$$

$$\mathbf{S}_{bci} = \mathbf{S}_{bcn} + \mathbf{S}_{wi}. \quad (3.6)$$

Az objektumon belüli szóródás hozzáadása ellenére, az SCDA módszer hátránya, hogy az osztályok közti szóródás nem pontos a helytelen normális eloszlás feltételezés miatt. Az alosztály diszkriminancia analízis módszerek azonban megoldást kínálnak erre a problémára, Mixture of Gaussians (MoG) eloszlás feltételezésével, és az alosztályok közti szóródás használatával. Ez a módszer azonban költséges (és a jelen esetben pontatlan) klaszterező lépésre alapul.

Szerencsére ez a probléma áthidalható a klaszterező eljárás finomításával (SDA-IC módszer). Ez az eljárás a klaszterek számát külön minden osztályhoz megbecsüli az osztályon belüli maximális méretű objektum alapján. Az alosztályok közepi ennek az objektumnak az elemei alapján kerülnek inicializálásra. Ezt követően a megmaradó csomópontokat egyszerűen a legközelebbi alosztályba soroljuk. Ez a módszer ekvivalens a hagyományos k-Means algoritmus egyetlen iterációjával. A hozzárendelés során lehetséges az algoritmust büntetni azért, ha egy objektum két csomópontja ugyanabba az alosztályba kerül.

Mivel az alosztályok inicializációját az objektumon belüli szeparációs kritérium figyelembevételével végeztük, így a kapott klaszterek közel lesznek az optimálishoz. A javított klaszterezés módszer így képes sokkal jobb klaszterezés elérésére, miközben nagymértékben csökkenti a szükséges számításokat.

Az SDA-IC eljárás apróbb hibája, hogy a két kritériumot ugyanabba a szóródási mátrixba vonja össze, ami lehetetlenné teszi azok relatív súlyának tervezői meghatározását. Ez pedig fontos lehet, ugyanis egyes alkalmazások az egyik fajta hibával szemben sokkal toleránsabbak lehetnek. Ez indokolhatja a mesterséges bias hozzáadását a dimenzió redukciós módszerhez. A megoldás erre a problémára az SCDA és az SDA-IC módszerek kombinálása az alosztály strukturált kompozit diszkriminancia analízis (SSCDA) módszerbe:

$$\mathbf{S}_{bci} = \mathbf{S}_{bsb} + \mathbf{S}_{wi}, \quad (3.7)$$

ahol \mathbf{S}_{bsb} az SDA módszer alosztályok közti szóródási mátrixa, ahol az alosztályok a javított klaszterezési módszerrel kerülnek meghatározásra.

A legjobban teljesítő módszer az SSCDA módszer lett, amely a rendelkezésre álló tesztadatbázisokon szignifikánsan felülmúlta a többi módszer pontosságát, az SCDA módszer kivételével, ahol a javulás kevésbé bizonyos ugyan, de még mindig valószínűleg pozitív. Az SCDA és SSCDA módszerek a 3D alak-gráf adatbázisokon is a legjobb teljesítményt érték el, így ezek jól használható módszerek bizonyultak az alakfelismerő rendszerhez.

Figyelemre méltó, hogy a megalkotott dimenzió redukciós módszerek nem korlátozottak egy konkrét alkalmazásra. Mint korábban említettük, strukturált adatok számos helyen előfordulnak az észlelés problémái között, például osztályozás és detekció esetében. A kísérleteink alapján az SSCDA módszer használható képjellemző leíróként, azonban - a SURF módszerrel ellentétben - nem feltétlenül invariáns képtranszformációkra, mint például a forgatás vagy a skálázás.

Általánosságban, az SSCDA eljárás alkalmazható részekből álló adatstruktúrákon végzett dimenzió redukcióra, amíg az egyes komponenseket ugyanazok a változók írják le. Továbbá a bemutatott módszerek feltételezték, hogy az osztályok MoG eloszlásúak, ami egy fontos limitációja ezeknek az eljárásoknak. Ha ez a feltételezés nem teljesül, az eljárás valószínűleg szuboptimális eredményt ad.

2. Tézis

Újszerű módszert alkottam vektorok halmazából vagy grájfjából álló (strukturált kompozit) osztályokon elvégzendő dimenzióredukció elvégzésére. Számos adatbázis segítségével megmutattam, hogy a módszer nem csak az egyes osztályok, hanem ez egyes objektumokon belüli elemek elválasztására is alkalmas. Statisztikai tesztekkel igazoltam, hogy a módszer strukturált kompozit osztályok esetén olyan leírókat szolgáltat, amelyek pontosabb osztályozást eredményeznek korábbi módszerekhez képest. [6, 7, 8]

3.3. Jelenetek berendezése

Ebben az alfejezetben az osztályozás eredménye alapján a globálisan optimális elrendezés meghatározására használt módszertant mutatjuk be. El a lépés több okból is szükséges: Egyrészt, az egyes részek optimális címkéi nem függetlenek egymástól. Mivel a legtöbb valódi objektum fizikailag egybetartozó, ezért az egymáshoz közeli részek nyilvánvalóan nagyobb eséllyel tartoznak ugyanahhoz az objektumhoz. Következésképp, az egyik rész végső címkéje befolyásolja a közeli részek optimális címkéjét.

Ezen felül az alkalmazás különböző egyéb megkötéseket is hozhat, amiket a végső elrendezésnek ki kell elégítenie. A tapintható kevert valóság esetben egyes objektumkategóriák jelenléte szükséges lehet, vagy a maximális számukon lehet korlát.

A módszer alapja egy objektív függvény konstruálása, amely az eredeti osztályozás közelében maradva ugyan, de az extra kritériumokat (pl.: az objektumok kompakt-sága) is figyelembe veszi. Továbbá a költségfüggvény flexibilitása is fontos szempont, ugyanis ez lehetővé teszi, hogy a környezet fejlesztője további saját megkötéseket vezessen be, például bizonyos objektumból minél több behelyezése. Végezetül, a költségfüggvénynek lehetővé kell tennie bizonyos kategóriapárok térbeli elrendezésének befolyásolását, valamilyen kontextus jutalom bevezetésével. Összegzésképp, az objektív függvénynek a következő elemeket kell magába foglalnia:

- Az elrendezés legyen megvalósítható: Minden csomópontnak pontosan egy címkéje legyen.
- A végső elrendezés osztályozási jósága legyen elfogadható.
- Közeli objektumrészeknek legyen megegyező címkéje.
- Minden szükséges kategóriához tartozzon legalább egy csomópont.
- Az algoritmust ösztönözzük, hogy minél többet helyezzen be olyan objektumokból, amelyeket a fejlesztő erre megjelölt.
- Az algoritmust ösztönözzük, hogy a 'kompatibilis' kategóriákat egymáshoz közel, míg az 'inkompatibilis' kategóriákat egymástól távol helyezze el.

A konstruált költségfüggvény négy komponensből tevődik össze: Az első a független osztályozásból származó jóságérték, míg a második a kompaktság tag, amely elősegíti a közeli csomópontok azonos címkéjét. A harmadik tag jutalmazza a helyes kontextust (kompatibilis kategóriák közelségét), míg a negyedik bizonyos kategóriák

darabszámát jutalmazza. Az objektumok szükségességét az optimalizálási problémába korlátozás formájában vezetjük be.

A költségfüggvény megalkotása után több különböző optimalizálási stratégia került kiértékelésre és összehasonlításra a problémán, beleértve egy mohó módszert, valamint a szimulált lágyítást és egy genetikus algoritmust. Ez utóbbi eljárás hatékonyságának növelése érdekében három újszerű genetikus operátort alkottam az inicializáció, a mutáció és a rekombináció műveletére.

A javasolt inicializációs módszer először az osztályozási jóságok alapján optimális megoldást adja hozzá a kezdeti populációhoz. Ezt követően, a mutációs operátor alkalmazásával legyártjuk a kezdeti populációt többi egyedét. A kezdeti populáció sokszínűségének biztosítása érdekében a duplum egyedeket újabb egyedi megoldásokkal helyettesítjük. Továbbá, a diverzitás növelése érdekében a kezdeti populáció bizonyos hányadát többszörös mutációk segítségével hozzuk létre.

A második megalkotott operátor a véletlen hurcoláson és keverésen alapul. A bináris/nominális integer genomok esetén a leggyakrabban a véletlen változtatás operátort alkalmazzák, ami egy véletlenszerű csomópont címkéjét változtatja meg. A jelen esetben azonban ez az operátor jó eséllyel szuboptimális egyedeket eredményez, ugyanis egyetlen csomópont címkéjének megváltoztatásával a kompaktság követelményt elrontjuk. Ezt oldja meg a véletlen hurcoláson: pont mutáció esetében nem csak a csomópont, hanem bizonyos eséllyel a közeli csomópontok címkéje is megváltozik. Ennek a valószínűsége állítja a pont és klaszter mutációk relatív gyakoriságát (látható, hogy ez előbbi szükséges az esetleges hibák javításához, vagy közeli objektumok szétválasztásához). Mindkét fajta mutáció engedélyezésével a paraméterteret hatékonyan tudjuk bejárni.

Az utolsó bevezetett operátor a klaszteres N -pont rekombináció. A hagyományos bináris/nominális integer rekombinációs operátor az N -pontos keresztezés. Ez az operátor behelyez N véletlen pontot a genomba, ezzel $N + 1$ szakaszra osztva azt. A leszármazott ezeket a szakaszokat felváltva öröklő a szüleitől. A probléma ezzel az operátorral a mutációs esethez hasonló, vagyis az ily módon történő öröklés nagy számú alacsony fitness értékű egyed létrehozásához vezetne.

Ezt a problémát a véletlen hurcoláshoz hasonló ötlet bevezetésével oldjuk meg: a keresztezést nem csomópontokon, hanem klasztereken definiáljuk. Ez azt jelenti, hogy a csomópontokat először a távolságuk alapján klaszterezzük egy adaptív küszöbérték segítségével. Ezt követően a klasztereket véletlenszerűen megkeverjük, és a N szakaszra osztjuk. A címkéket a leszármazott ezt követően szakaszonként felváltva öröklő a szülőktől.

A módszereket számos adatbázison kiértékelve az eredmények tisztán mutatják a globális optimalizálás eredményességét az egyszerű osztályozással szemben. Továbbá a saját operátorokkal kiegészített genetikus algoritmus vezetett magasban a legjobb eredményre, megmutatható, hogy mindhárom operátor *szignifikánsan* és pozitív előjellel járul hozzá a végső pontossághoz. Ezen felül az is belátható, hogy a beágyazási módszer javítja a végső elrendezés pontosságát. Végezetül a neurális háló alapú baseline-nal történő összevetés megmutatta, hogy a globális optimalizálás jobb eredményt ér el egy jelenet adatbázisokon tanított neurális hálónál.

A fizikai objektum térbeli kompaktságának kihasználása egyaránt kulcsfontosságú része volt a költségfüggvénynek és a genetikus operátoroknak, ami azt jelenti, hogy a módszerek teljesítménye e feltételezések igazságától nagymértékben függ. Habár elméletileg ez egy nyilvánvaló limitáció, azonban a logikai entitások tér- vagy időbeli kompaktsága meglehetősen univerzális jelenség. Ez persze nem jelenti, hogy nem léteznek kivételek, csupán azt, hogy ezek véleményünk szerint ritkán fordulnak elő.

3. Tézis

Átfogó és hatékony módszert alkottam globálisan optimális objektumelrendezések keresésére jelentekben, a probléma geometriai tulajdonságainak kihasználásával. A probléma megfogalmazása lehetővé teszi, hogy az optimalizálási módszer figyelembe vegye a Tapintható Kevert Valóság rendszerekben felmerülő tervezői szempontokat is. Nagyszámú jelenet-gráf adatbázison végzett statisztikai tesztekkel igazoltam, hogy az új algoritmus teljesítménybeli javulást eredményez korábbi gráf alapú objektumdetektáló módszerekhez képest.

(a) Újszerű genetikus operátorokat alkottam, amelyek statisztikailag szignifikánsan növelik a globális optimum megtalálásának esélyét olyan optimalizálási problémák esetében, ahol az objektumok kompaktsága feltételezhető.

[1, 2, 5, 9]

4 Eredmények összegzése

Az értekezés célja egy új módszerekből álló eljárás megalkotása volt, amely hatékonyan képes virtuális objektumok globálisan optimális elrendezését meghatározni egy 3D jelenetben. A párosítás 3D formabeli hasonlóságra alapszik, ami kiegészül számos további kritériummal, ami egy Adaptív Kevert Valóság (AKV) rendszer fejlesztője számára hasznos lehet. Ezek a kritériumok a virtuális objektumok behelyezését hivatottak segíteni, hogy a végső elrendezés intuitív és megfelelő legyen.

Az értekezésben három új megoldást ismertettünk, az első gráf csomópontok és kontextusuk vektortérbe történő beágyazását végzi, a második strukturált kompozit osztályokon végez diszkriminancia analízist, a harmadik pedig az objektumok globálisan optimális elrendezését határozza meg. Ezek az algoritmusok hatékonyak bizonyultak a fent leírt probléma megoldására, de nem korlátoztak egy specifikus alkalmazásra.

A javasolt módszerek eredményeit számos különböző adathalmazon értékeltem ki, az eredmények alapján alacsony hibaarányt megfelelő sebesség mellett sikerült elérni. Mivel a kezdeti elrendezést csak egyszer, a használat elején kell meghatározni, néhány másodperces futási idő elfogadható. A kísérleti eredmények alapján a javasolt módszerek pontosságbeli javulást eredményeznek korábbi irodalmi módszerekkel szemben.

Összefoglalásképp a megalkotott módszerek használható megoldást adnak a virtuális objektum AKV rendszerben történő elrendezésének problémájára. Említésre méltó, hogy habár a megoldás hibaaránya alacsony (a valós objektumokból álló adatbázison 0.62%), mégis eredményez hibás hozzárendelést elvéve. Ezek a hibák azonban legrosszabb esetben is kényelmetlenséget okoznak a felhasználónak, mivel a hibák manuálisan javíthatók. A hozzárendelés nélkül azonban az összes párosítást manuálisan kellene végezni. A reprodukálhatóság érdekében az algoritmusok és a felhasznált adatbázisok elérhetők online a <https://www.github.com/szemenyeim/PhD-Research> címen.

Publikációk

- [1] M. Szemenyei, „Neural Graph Node Classification via Self-Attention”, *Workshop on the Advances of Information Technology*, B. Kiss és L. Szirmay-Kalos, szerk., BME-IIT, 2020, 33–38. old.
- [2] M. Szemenyei és F. Vajda, „3D Object Detection and Scene Optimization for Tangible Augmented Reality”, *Periodica Polytechnica Electrical Engineering and Computer Science*, 62. évf., 2. sz., 25–37. old., 2018. DOI: 10.3311/ppee.10482.
- [3] M. Szemenyei és F. Vajda, „Learning 3D Object Recognition Using Graphs Based on Primitive Shapes”, *Workshop on the Advances of Information Technology*, B. Kiss és L. Szirmay-Kalos, szerk., BME-IIT, 2015, 67–71. old.
- [4] M. Szemenyei és F. Vajda, „3D Object Detection Using Vectorial Graph Node Embedding”, *Workshop on the Advances of Information Technology*, B. Kiss és L. Szirmay-Kalos, szerk., BME-IIT, 2017, 45–53. old.
- [5] M. Szemenyei és P. Reizinger, „Attention-Based Curiosity in Multi-Agent Reinforcement Learning Environments”, *2019 International Conference on Control, Artificial Intelligence, Robotics & Optimization (ICCAIRO)*, IEEE, 2019. máj. DOI: 10.1109/iccairo47923.2019.00035.
- [6] M. Szemenyei és F. Vajda, „Dimension Reduction for Structured Composite Classes in Multi-Object Environments”, *Proceedings of the 15th International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases (AIKED)*, V. Mladenov, szerk., Recent Advances in Electrical Engineering sor., WSEAS, 58. köt., WSEAS Press, 2016, 134–141. old.
- [7] M. Szemenyei és F. Vajda, „Dimension Reduction for Objects Composed of Vector Sets”, *International Journal of Applied Mathematics and Computer Science*, 27. évf., 1. sz., 169–180. old., 2017. DOI: 10.1515/amcs-2017-0012.
- [8] M. Szemenyei és F. Vajda, „Optimal Feature Selection for Objects Composed of Vector Sets”, *Hungarian Conference on Computer Graphics and Geometry*, L. Szirmay-Kalos és G. Renner, szerk., NJSZT, 2016, 7–14. old.
- [9] M. Szemenyei, „Evolutionary Scene Arrangement for Adaptive Augmented Reality Systems”, *Workshop on the Advances of Information Technology*, B. Kiss és L. Szirmay-Kalos, szerk., BME-IIT, 2018, 10–17. old.