

VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
INFORMATIKOS INSTITUTAS
INFORMATIKOS KATEDRA

Generatyviniai besivaržantys neuroniniai tinklai 3D modeliams generuoti iš 2D nuotraukų

**Generative Adversarial Networks for 3D Model Reconstruction
from 2D Images**

Magistro baigiamojo darbo literatūros apžvalga

Atliko:	Marius Bieliauskas	(parašas)
Darbo vadovas:	dr. Olga Kurasova	(parašas)
Recenzentas:	dr. Aistis Raudys	(parašas)

Turinys

Santrauka	2
Summary	3
1. Įvadas	4
2. Literatūros apžvalga	6
2.1. 3D skaitmeninimas ir fotogrametrija	6
2.1.1. Taškinių debesų generavimo algoritmai	7
2.1.2. Taškinių debesų apdorojimo algoritmai	11
2.2. Generatyviniai besivaržantys tinklai	11
2.3. 3D GAN	14
2.3.1. IKEA baldų skaitmeninimas	14
2.3.2. Archeologinis paveldas	15
2.4. 3D modelių kokybės tobulinimas	18
3. Išvada	21
Literatūra	22

Santrauka

Atliekama literatūros apžvalga siekiant surinkti visą aktualią informaciją dirbtinio intelekto metodo kūrimui, kuris, naudojant generatyvinius besivaržančius tinklus (angl. Generative adversarial networks arba GAN), galėtų iš vienos ar kelių dvimačių nuotraukų sukonstruoti trimatę fotografuoto objekto reprezentaciją. Apžvelgiami tradiciniai trimacio skaitmeninimo metodai, bendra GAN teorija ir įvairūs GAN metodų patobulinimai, esami darbai atliekantys trimatių skaitmeninimą su GAN pagalbą bei 3D modelių išlyginimo ir kokybės gerinimo metodai. Literatūros analizės metu siekiama identifikuoti šiuo metu taikomus metodus ir jų efektyvumą bei rasti galimas jų tobulinimo kryptis.

Raktiniai žodžiai: Dirbtinis intelektas, GAN, fotogrametrija, 3D skaitmeninimas.

Summary

A literature review is performed in order to collect all relevant information for the creation of an artificial intelligence based method which, using generative adversarial networks (GAN), will be able to construct three-dimensional representations of objects captured in one or more photographs. Traditional three-dimensional digitization methods, general GAN theory and various improvements to GAN methods, current work on three-dimensional digitization with the help of GAN, and methods for smoothing and improving the quality of 3D models are reviewed. The aim of the literature review is to identify the currently applied methods, their performance and find possible avenues for their improvement.

Keywords: Artificial intelligence, GAN, photogrammetry, 3D digitization.

1. Įvadas

Magistro baigiamajam darbui pasirinkta tema "Generatyviniai besivaržantys neuroniniai tinklai 3D modeliam generuoti iš 2D nuotraukų". Šio darbo metu bus nagrinėjami dirbtinio intelekto metodai, konkrečiai generatyviniai besivaržantys neuroniniai tinklai, toliau GAN (angl. Generative Adversarial Networks), norint išgauti trimatę objektą, esančią vienoje ar keliose dvimatėse nuotraukose, reprezentaciją.

Dirbtinio intelekto ir tiksliau gilaus mokymo srityje daugiausiai sėkmės yra susilaukę diskriminuojantys (angl. discriminative) modeliai, kurie yra naudojami klasių priskyrimui prie gautų duomenų, tačiau generatyviniai modeliai, kurie gali sukurti naują informaciją iš turimų duomenų, ilgą laiką atsiliko savo kokybe. GAN tinklai buvo prasiveržimas generatyvių modelių srityje, ypač užduotyse susijusiuose su vaizdine informacija. Nuo pirmojo GAN pasiūlymo 2014 metais Ian J. Goodfellow ir kt. parašytame straipsnyje tuo pačiu pavadinimu [GPM⁺14], ši metodologiją labai greitai tobulėjo ir privėdė prie itin gerų rezultatų kuriant naują informaciją, kuri yra sunkiai atskiriama nuo realios. Vieni labiausiai stebinančių rezultatų gauti žmonių veidų generavimo srityje. 2017 metais Tero Karras ir kt. parašytame straipsnyje [KAL⁺18] gauti fotorealistički neegzistuojančių žmonių veidai, be jokių įrodymų, kad modelis tiesiog atsiminė ir atkartojo jau egzistuojančius veidus iš mokymo duomenų. GAN tobulėjimą šioje srityje galima pamatyti 4 paveikslėlyje.

GAN tinklai yra viena naujausių ir aktualiausių dirbtinio intelekto technologijų, dirbanti su vaizdine medžiaga. Vienas mažiau išnagrinėtų GAN pritaikymų yra fotogrametrijos ir trimačių modelių kūrimo (skaitmeninimo) srityje. 3D modeliavimo sujungimas su GAN technologijomis aprašomas JiaJun Wu ir kt. straipsnyje [WZX⁺16] bei V.V. Kniaz ir kt. straipsnyje [KRK19]. Šis sujungimas gali būti pritaikomas kultūrinio paveldo atkūrimui, pramogų ir ypač kompiuterinių žaidimų srityse, bei kituose srityse, kuriose svarbus objektų 3D skaitmeninimas. Jeigu 3D modeliavime galėtų būti matomas panašus tobulėjimas kaip ir kitose GAN srityse, šios technologijos ženkliai sumažintų dabartinius kaštus visose anksčiau minėtose srityse, kadangi šiuo metu geriausios 3D skaitmeninimo technologijos reikalauja brangių gylio matavimo ar lazerinio skenavimo įrankių, norint išgauti reikiama kiekį informacijos.

Magistro baigamojo darbo tikslas yra pasiūlyti metodą trimačių modelių išgavimui iš dvimačių nuotraukų, naudojant generatyvinius besivaržančius neuroninius tinklus, kuris būtų pranašesnis lyginant su esamais metodais. Siekiant darbo tikslo bus atliekami šie uždaviniai:

1. Igyvendinti veikiantį GAN tinklą, kuris galėtų iš dvimačių nuotraukų išgauti fotografuojamų objektų trimates reprezentacijas.
2. Optimizuoti igyvendintus neuroninius tinklus, kad gauti kiek įmanoma geresnius rezultatus tiktai su viena nuotrauka.
3. Pridėti galimybę naudoti kelias nuotraukas vieno objekto modelio sukūrimui, siekiant, kad panaudojus mažiau nuotraukų būtų gaunami geresni rezultatai nei su tradiciniais daugelio nuotraukų skaitmeninimo metodais.

4. Pritaikyti automatinio išlyginimo (angl. smoothing) algoritmą, kuris galėtų patobulinti gautų 3D modelių raišką ir kokybę.

Šiame konkrečiame darbe pateikiama pirmoji magistro baigamojo darbo dalis – aktualios literatūros apžvalga. Apžvelgiami tradiciniai trimačio skaitmeninimo metodai, bendra GAN teorija ir įvairūs GAN metodų patobulinimai, esami darbai atliekantys trimatį skaitmeninimą su GAN pagalbą bei 3D modelių išlyginimo ir kokybės gerinimo metodai. Šio darbo tikslas – suprasti šiuo metu taikomus metodus ir jų efektyvumą bei rasti galimas jų tobulinimo kryptis.

2. Literatūros apžvalga

Norint pasiekti šio magistro baigiamojo darbo tikslą – pasiūlyti trimačių modelių išgavimo iš dvimačių nuotraukų metodą, naudojant besivaržančius neuroninius tinklus, kuris būtų pranašenės lyginant su esamais metodais – buvo atlikta literatūros analizė, siekiant suprasti šiuo metu tam naudojamas technologijas ir metodus, jų efektyvumą ir galimas tobulinimo galimybes. Literatūros apžvalga apims aktualią fotogrametrijos ir trimačio skaitmeninimo teoriją, esamus metodus ir įgyvendinimus, jų pasiekimus ir trukūmus, bei GAN technologijas ir jų pritaikymą fotogrametrijos rezultatų tobulinimui. Taip pat apžvelgiami trimačių modelių išlyginimo (angl. smoothing) ir raiškos padidinimo (angl. upscaling) algoritmai, kurie galėtų būti pritaikyti galutinio skaitmeninto 3D modelio kokybei pagerinti.

2.1. 3D skaitmeninimas ir fotogrametrija

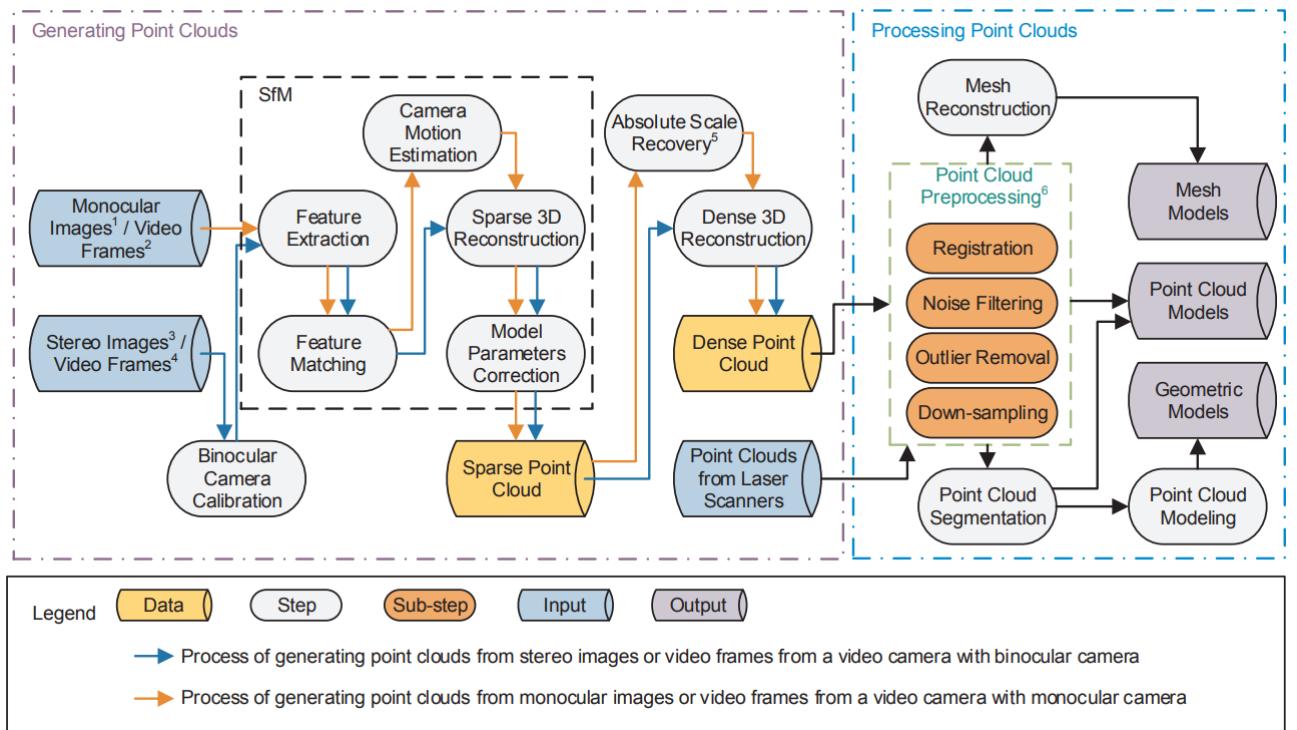
Vienas pagrindinių šio magistro baigiamojo darbo rezultatų turės būti suskaitmenintas trimačis modelis, sukurtas pagal realaus objekto atvaizdą. Trimatis modelis – tai skaitmeninis objekto atvaizdavimas sudarytas iš rinkinio (tinklo) sujungtų taškų trimatėje erdvėje. Šie modeliai dažniausiai naudojami trimatei kompiuterinei grafikai kompiuteriniuose žaidimuose, filmuose, architektūroje, interjero dizaine, kultūrinio paveldo išsaugojimui bei analizei, ir kt.

3D modeliavimas nagrinėjamas jau ilgą laiką, nuo pat jo atsiradimo praeito amžiaus aštuntame dešimtmetyje. Savaime suprantama, kad per tą laiko tarpą šia tema buvo pasiūlyta daugelis skirtinės metodų ir algoritmų. Modelių kūrimas pagal realius objektus, neskaitant rankinio atkartojimo, gali būti skaidomas į dvi dalis – fotogrametrinius metodus ir skenavimą (dažniausiai lazerinį ske navimą).

Fotogrametrija – tai mokslo šaka, tirianti objekto formos, matmenų ir (arba) padėties nustatyti iš fotografinių nuotraukų. Kai kalbama apie fotogrametriją, dažnai kalbama apie tradicinę iš oro arba iš orbitos gautą žemės paviršiaus nuotrauką tyrimą, tačiau šiam darbui pritaikoma artimojo nuotolio fotogrametrija, o tiksliau trimačių modelių išgavimas iš nuotraukų. Vienas standartinių fotogrametrijos metodų pritaikymo pavyzdys matomas Jūratės Sužiedelytės-Visockienės ir kt. 2015 metų straipsnyje [SBM⁺15] kuriame bandoma išgauti trimatę informaciją apie architektūros paveldo deformacijas. Straipsnyje minima, kad nors su lazeriniais skenavimais gaunami geresni trimačio skaitmeninimo rezultatai, dažnai lazerinio skenavimo galimybės nėra ir turi būti atliekama nuotraukų analizė.

Trimatis skenavimas – kitaip nei fotogrametrijoje, vietoje nuotraukų yra naudojami specializuoti prietaisai skirti atstumų nustatymui trimatėje erdvėje. Dažniausiai šiam tikslui naudojami lazeriniai pagrįsti skaitytuviniai, tačiau tam tikrose srityse naudojamos ir kitos technologijos, kaip kompiuterinė ar viršgarsinė tomografija.

2018 metais Zhiliang Ma ir Shilong Liu atliko tuo metu prieinamų trimačio skaitmeninimo (rekonstrukcijos) metodų civilinės inžinerijos srityje [ML18]. Autoriai iš viso išanalizavo 95 skirtinės straipsnius, iš kurių 88, kaip įvesti naudojantys monofoninių ir stereofoninių kamerų vaizdus arba vaizdo įrašus bei lazerinių skaitytuvų išvesti, buvo aprašomi šioje apžvalgoje. Jų parašytame



1 pav. Trimačio skaitmeninimo žingsniai [ML18]. Pastabos: 1 – vaizdai gauti iš monofoninių kamerų, 2 – vaizdo įrašai gauti iš monofoninių kamerų, 3 – vaizdai gauti iš stereofoninių kamerų, 4 – vaizdo įrašai gauti iš stereofoninių kamerų.

straipsnyje galima rasti glauštą aprašymą apie bendrą skaitmeninimo procesą bei tam naudojamus algoritmus.

Autoriai dalina trimatį skaitmeninimą į du žingsnius: taškinių debesų generavimą ir jų apdorojimą. Pirmajame žingsnyje, taškiniai debesys yra generuojami iš vaizdinės medžiagos, bandant iš daugelio skirtinčių kampų ir atstumais paimtų nuotraukų išgauti taškų, patalpintų ant viso atvaizduojamo objekto paviršiaus, koordinacijų rinkinį. Šis žingsnis nėra pritaikomas naudojant lazerinius skaitytuvus, kadangi šių prietaisų išvestis, nuskenavus objektą, ir yra norimas taškų debesis. Antrajame žingsnyje šis taškų rinkinys yra apdorojamas, tam kad sukurti norimą trimatėlio modelio išvesti. Apibendrintą objekto trimatėlio skaitmeninimo procesą galima matyti 1 paveikslėlyje.

2.1.1. Taškinių debesų generavimo algoritmai

Ma ir Liu savo apžvalgoje [ML18] rado, kad taškinių debesų generavimo žingsnis susideda iš septynių mažesnių, iš eilės atliekamų žingsnių: ypatybių išgavimas (angl. feature extraction), ypatybių prilyginimas (angl. feature matching), kameros jūdesio įvertinimas (angl. camera motion estimation), reta 3D rekonstrukcija (angl. sparse 3D reconstruction), modelio parametru koregavimas (angl. model parameters correction), absoliutaus (tikrojo) mastelio atkūrimas (angl. absolute scale recovery) ir galiausiai tanki 3D rekonstrukcija (angl. dense 3D reconstruction). Pirmieji penki šio proceso žingsniai dar bendrai vadinami struktūra iš jūdesio, arba SfM (angl. Structure from Motion).

Ypatybių išgavimo tikslas yra surinkti ypatybių taškus, kurie atspindi pradinę skaitmeninamo objekto struktūrą, iš vaizdinės medžiagos. Šiame žingsnyje yra naudojami ypatybių taškų detekto-

riai, kurie randa šią taškų poziciją fotografuojamoje ar filmuojamoje scenoje, ir taškų aprašytojai (angl. descriptors), kurie aprašo taškų savybes. Autorių apžvelgtuose straipsniuose rasti tokie detektoriaus metodai ir jų veikimo principai:

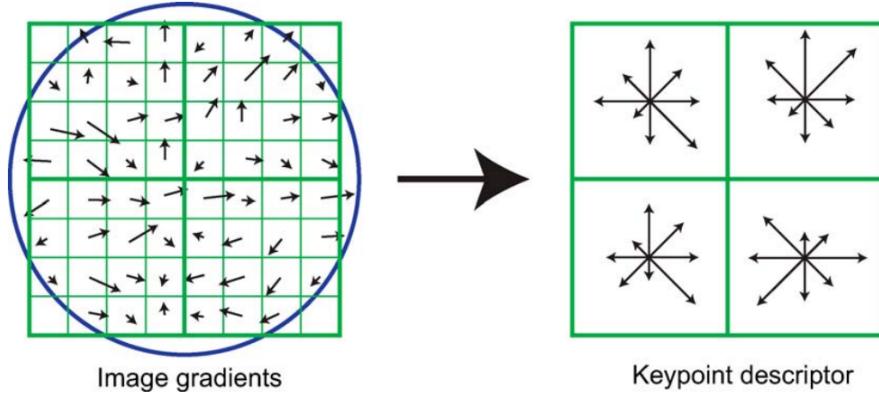
- SIFT (angl. Scale Invariant Feature Trasform) detektorius [Low04] – šiame metode iš pradinių nuotraukų yra sugeneruojami įvairaus mastelio vaizdų rinkiniai – mastelio erdvės (angl. scale-space), pritaikant konvoliuciją su Gauso funkcija, kuri matoma formulėje (1) (x ir y formulėje atitinka pikselių koordinates). Tada juos vieną iš kito atimant gaunami Gauso skirtumo vaizdai (angl. difference-of-Gaussian images), kuriuose kiekvienas pikselis yra lyginamas su kaimyniniais pikseliais ir ieškoma ekstremumo taškų, kurie vėliau tampa ypatybių taškais.

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (1)$$

- SURF (angl. Speeded-Up Robust Features) detektorius [BET⁺08] – vėlgi generuojama mastelio erdvė tačiau vietoje originalių nuotraukų yra naudojami jų vientisi vaizdai (angl. integral images), kurie kiekvienoje vaizdo koordinatėje laiko sumą visų pikselių verčių, esančių stačiakampyne į viršų ir į kairę nuo tos koordinatės. Tai leidžia vietoje Gauso funkcijos naudoti kvadratinius filtrus, kadangi pikselių verčių sumavimas tam tikroje vaizdo dalyje reikalauja žinoti tik tai kampų vertes ir yra skaičiuojamas daug greičiau. Taip pat vietoje Gauso skirtumo yra skaičiuojamas Hesiano determinantas kiekvienam vaizdui mastelio erdvėje ir pagal jį randami ypatybių taškai.
- ASIFT (angl. affine-SIFT [MY09] – tai SIFT metodo papildymas, kuriame prie mastelio erdvės yra pridedami sugeneruoti vaizdai, imituojančiais kameros pasukimą.
- Harris kampų detektorius [HS88] – tai paprastesnis algoritmas, kuris bando rasti nuotraukose esančių objektų kampus (kraštus) stebint staigius pikselių verčių pasikeitimus.
- FAST (angl. Features from Accelerated Segment Test) [RD06] – kampų radimo metoda, kuris greitesnis už Harris algoritmą. Jis kiekvienam pikseliui analizuojant 16 kaimyninių pikselių, išsidėsčiusių apskritimu nuo jo, ir jeigu tam tikra dalis jų peržengia tam tikrą šviesumo slenkstę šis pikselis yra laikomas kampo dalimi.

Deskriptoriai:

- SIFT aprašytojas [Low04] – apart ypatybių aptikimo, SIFT metodas taip pat susideda ir iš ypatybių aprašymo žingsnio. Jame vektoriai aprašantys ypatybių taškus yra generuojami remiantis vaizdo taškų, esančių aplink ypatybės tašką, gradiento dydžiais ir kryptimis. Vizualų to pavaizdavimą galima matyti 2 paveikslėlyje.

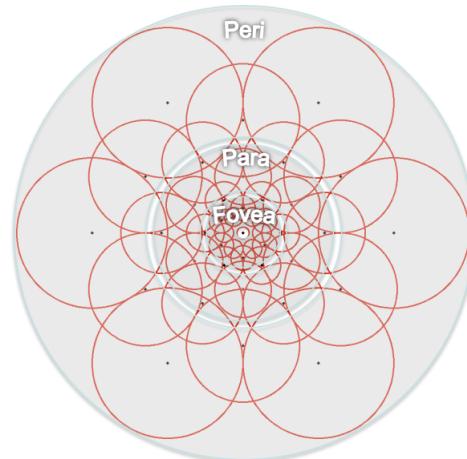


2 pav. Ypatybė aprašančio vektoriaus sukūrimas sujungiant aplink ypatybės tašką esančių pikselių gradientus SIFT metode [Low04].

- SURF aprašytojas [BET⁺08] – šiame metode ypatybių taškai aparašomi skaičiuojant Haar bangelių reikšmes x ir y koordinatėms (formulė (2)), pikseliams tam tikru spinduliu aplink juos.

$$\varphi(t) = \begin{cases} 1 & \text{Jei, } 0 \leq t < \frac{1}{2} \\ 0 & \text{Jei, } \frac{1}{2} \leq t \leq 1 \end{cases} \quad (2)$$

- FREAK (angl. Fast REtinA Keypoint) [AOV12] – šis metodas renka apskritime aplink ypatybės tašką esančius pikselius žmogaus akies tinklainė pagrįstu metodu, t.y. imamos mažesnės grupės esant arčiau ypatybių taško ir didesnės grupės esant toliau. Taip taškams esantiems arčiau ypatybės taško yra suteikiama didesnė vertė (svoris) nei taškams esantiems toliau. Šio metodo rezultate gautas aprašas yra dvejetainė eilutė apskaičiuota atliekant vieno bito Gauso skirtumo operacijas tarp rinkinių porų. Kaip gaunami taškų rinkiniai galima pamatyti 3 paveikslėlyje.



3 pav. Pikselių rinkimas aplink ypatybės tašką FREAK metode. Visi apskritimai yra imlūs laukai (angl. receptive fields) pikselių, pagal kurios bus skaičiuojama ypatybės tašką aprašanti dvejetainė eilutė. [AOV12].

- MSD (angl. Multi-Scale Descriptor) [SC13] – aprašymo generavimas šiame metode pagrindžiamas pikselių, esančių aplink ypatybės tašką, šviesumo (angl. intensity) gradientais, keliuose skirtinguose masteliuose.

Ypatybių prilyginimo algoritmai bando rasti sutapimus tarp gautų ypatybių taškų lyginant nuotraukų poras. Tam naudojami apytikslių arčiausią kaimynų algoritmai, toliau ANN (angl. Approximate Nearest Neighbours), kurie skaičiuoja Euklido atstumus tarp ypatybių taškų aprašų. Kad panaikinti klaudingus atitikimus ir išlaikyti objekto geometrinę informaciją, yra naudojami RANSAC [FB81] arba ORSA [RGL⁺14] algoritmai, imantys nuotraukų poras ir pažingsniui skaičiuojantys scenos epipolinę geometrinę informaciją (Epipolinė geometrija – tai stereo matymo geometrija, kai dvi kameros mato tą pačią trimatę sceną iš dviejų skirtinį padėcių).

Kameros jūsudėsio įvertinimo tikslas yra kiekvienai nuotraukai surasti ją paveikiančius kameros parametrus. Šie parametrai gali būti vidiniai, kaip kameros objektyvo židinio nuotolis ir radialinis iškraipymas arba tai gali būti išoriniai parametrai, kaip kameros pasukimas, pajudinimas ar pakreipimas. Kai turima nuotraukų pora, dvi matricos gali nusakyti jų epipolinės geometrijos informaciją – esminė matrica (angl. essential matrix), kuri laiko tiktais išorinius parametrus, ir pagrindinė matrica (angl. fundamental matrix), kuri laiko ir vidinius ir išorinius parametrus. Jų radimui atitinkamai yra naudojami du, penkių taškų [Nis04] ir aštuonių taškų [Har97], algoritmai, kurie pagal savo pavadinimus paima arba penkis arba aštuonis prilygintus taškus nuotraukose ir pagal juos skaičiuoja vieną ar kitą matricą. Kadangi esminė matrica nelaiko vidinių kameros savybių, ji ir penkių taškų algoritmas naudojami tiktais tada, kai kameros vidinės savybės jau yra žinomos ir nereikia švaistyti laiko jų ieškant.

Sekantis žingsnis, reta 3D rekonstrukcija, siekia apskaičiuoti taškų padėtis trimatėje erdvėje ir gauti fotografuojamą sceną ar objektą atitinkantį taškų debesį, naudojant praeituose žingsniuose gautus ypatybių taškų tarpusavio atitikimus ir kiekvienos nuotraukos kameros parametrus. Tam naudojamas triangulacijos algoritmas ([LCS⁺16], [YCH⁺13], [KL17]).

Turint retą scenos taškinį debesį yra atliekamas kameros parametru koregavimas. Šiame žingsnyje bandoma ištaisyti ir sulyginti visų nuotraukų parametrus. Tam yra naudojamas ryšulio reguliavimo algoritmas (angl. Bundle adjustment), pagrįstas netiesinių mažiausią kvadratų metodu [TMH⁺00].

Prieš atliekant galutinio taškų debesies išgavimo žingsnį yra randamas absolitusis (tikrasis) objekto mastelis, sulyginant realias koordinates ir atstumus tarp įvairių taškų su gautais praeituose žingsniuose. Tam gali būti naudojamas geografinės registracijos (angl. geo-registration arba geo-referencing) algoritmas [ZHK⁺14] kuris apskaičiuoja nuotraukų pasukimo ir perkėlimo matricas tarp turimo taškų debesies ir tikrosios koordinacių sistemos ir jomis transformuoja taškų debesį. Taip pat galima mastelio suvienodinimą atlikti rankiniu būdu, pamatuojant atstumus tarp vienos ar kelių porų taškų ant realaus objekto, ir pagal tai išsivedant visą 3D modelio mastelį. Tačiau naudojant tokį metodą labai lengvai gali atsirasti klaidų, taip pat jis ganėtinai lėtesnis.

Galiausiai atliekama tanki 3D rekonstrukcija, papildant turimą retą taškų debesį su informacija iš nuotraukų su pataisytais kameros vidiniais ir išoriniais parametrais. Tam yra naudojami PMVS [FP10] arba SGM [Hir05] metodai, bendrai paėmus, autoriai rado, kad SMG algoritmas greitesnis,

o PMVS yra tikslėsnis ir stabilesnis.

Ma ir Liu straipsnyje taip pat išskiriami žingsnių pokyčiai stereofoninėms kameroms (galima pamatyti 1 paveikslėlyje) ir vaizdo įrašams, tačiau šiam darbui tai nėra aktualu.

2.1.2. Taškinių debesų apdorojimo algoritmai

Dažniausiai taškiniai debesys gauti iš aukščiau aprašytų žingsnių ar iš lazerinių skaitytuvu neatitinka modelių pritaikymo reikalavimų, kaip nereikalingų elementų iš įsiterpimas iš scenos fono. Tad norint atitikti šiuos reikalavimus turimi taškiniai debesys yra apdorojami atliekant dar keturis žingsnius: taškinio debesies išankstinis apdorojimas (angl. preprocessing), tinklelio rekonstrukcija (angl. mesh reconstruction), taškinio debesies segmentacija ir galiausiai jo modeliavimas. Tam naudojamus algoritmus galima rasti 1 lentelę.

1 lentelė. Ma ir Liu straipsnyje [ML18] apžvelgti taškinių debesų apdorojimo metodai ir algoritmai.

Žingsnis	Požingsnai	Algoritmai ir metodai
Taškinio debesies išankstinis apdorojimas	registracija triukšmo filtravimas išskirčių pašalinimas raiškos sumažinimas (angl. downsampling)	ICP (angl. Iterative closest point) [CSS ⁺ 02] Rankinis taškų pašalinimas RANSAC [FB81] Taškų atstumo strategija (angl. Point spacing strategy) [SKK15]
Tinklelio rekonstrukcija	-	PSR (angl. Poisson surface reconstruction) [Hop08]
Taškinio debesies segmentacija	-	Regiono auginimas (angl. Region growing) K-vidurkių klasterizavimas (angl. K-means clustering) Vokseliai (angl. voxel) pagrįsti algoritmai Hough transformacija RANSAC [FB81]
taškinio debesies modeliavimas	-	Objektų matmenų išgavimas [PHL ⁺ 17]

2.2. Generatyviniai besivaržantys tinklai

Generatyviniai besivaržantys neuroniniai tinklai, toliau GAN (angl. Generative Adversarial Networks), buvo pirmą kartą pasiūlyti 2014 metais Ian J. Goodfellow ir kt. straipsnyje tuo pačiu pavadinimu [GPM⁺14]. Naudojant šį metodą vienu metu yra mokomi du neuroninių tinklų modeliai, kurie tarpusavyje konkuruoja. Pirmasis modelis yra generatyvinis ir bando pagal naudotus mokymo duomenis sukurti tokią išvestį, kuri apgautų kitą, diskriminuojantį modelį, kuris gavęs generatyvinio modelio rezultatus, turi atskirti, ar jie priklauso realiems duomenims, ar jie yra sugeneruoti. Taip gaunamas modelių tarpusavio nulinės sumos žaidimas, kuriame vieno modelio sėkmė yra kito modelio nesėkmė, t.y. jeigu generatyvinio modelio rezultatai pagerėja ir jis dažniau apgauna diskriminuojantį modelį, tai reiškia, kad diskriminuojančio modelio rezultatai proporcingai suprastėja (ir atvirkščiai), tad bendrai abiejų modelių rezultatų suma išlieka nulinė.

Paprasčiausias šio metodo pritaikymas gaunamas kai abu neuroninių tinklų modeliai yra daugiasluoksniai perceptronai. Jie gali būti išreiškiami kaip $G(z; \theta_g)$ ir $D(x; \theta_d)$. G – tai generatyvinio

modelio funkcija, kuri iš atsitiktinio triukšmo z generuoja naują informaciją naudojant parametrus θ_g . D – tai diskriminuojančio modelio funkcija išvedanti vieną skaitinę reikšmę, kuri nusako tikimybę, kad duomenų rinkinio elementas x yra dalis originalių duomenų, o ne G modelio sugeneruotų duomenų, naudojant parametrus θ_d .

Modelių apmokymo metu yra atliekamas savotiškas minimax žaidimas, kuris naudojamas ir kituose nulinės sumos žaidimuose kaip šachmatai. Jo eigoje modelis D yra mokomas, kad maksimizuoti tikimybę teisingai nuspėti ar duomenys yra realūs ar sugeneruoti, o tuo pačiu modelis G yra mokomas minimizuoti tą pačią tikimybę. Tai išreiškiama su šia netekties funkcija (angl. loss function), kurioje $V(D, G)$ yra žaidimo vertės funkcija (angl. value function):

$$\min_G \max_D V(D, G) = \mathbb{E}_x[\log D(x)] + \mathbb{E}_z[\log(1 - D(G(z)))] \quad (3)$$

Autoriai rado, kad praktikoje šis žaidimas turi būti įgyvendinamas pažingsniui, kadangi vienas ar kitas modelis gali tapti permokytas. Tam siūloma atlikti keli D apmokymo žingsnius prieš atliekant vieną G mokymo žingsnį. Kita problema yra prasti gradientai G mokymo pradžioje, kadangi D labai lengvai atskiria sugeneruotus duomenis, kad to išvengti galima vietoje $\log(1 - D(G(z)))$ minimizavimo, maksimizuoti $D(G(z))$, taip išreiškiamas tas pats tikslas, tačiau gaunami daug reikšmingesni gradientai.

Šio žaidimo visuotinis optimumas gaunamas, kai diskriminuojančio modelio tikimybė pasirinkti sugeneruotus duomenis ir pasirinkti originalius duomenis tampa vienoda. Tai reiškia, kad algoritmas nebesugeba atskirti sugeneruotų duomenų nuo realių, t.y tikimybių pasiskirstymai lygūs – $p_g = p_{\text{data}}$ ir $D(x) = \frac{1}{2}$.

Šie tinklai buvo pritaikyti daugelyje skirtinų dalykinių sričių, tačiau daugiausiai sekmės susilaukė vaizdinės medžiagos apdorojimo ir kūrimo srityje – [ZXL⁺17], [IZZ⁺17] ir [LTH⁺17]. Po pirmojo Goodfellow GAN paskelbimo taip pat buvo pasiūlyta daugelis patobulinimų. Keli jų aprašyti 2016 metų T. Salimans ir kt. straipsnyje [SGZ⁺16]. Jame yra išsikeliamas GAN konvergavimo problema. Šie tinklai paprastai naudoja gradientinio nusileidimo metodus, ieškančius žemos netekties funkcijos reikšmės, tačiau šie metodai gali nesugebėti konverguoti (suartēti). Tai atsitinka kai vienas iš modelių pradeda dominuoti ir abiejų modelių rezultatai nustoja tobuleti.

Salimans straipsnyje yra pasiūlomi penki konvergavimo skatinimo būdai:

1. Ypatybių prilyginimas (angl. Feature matching) – bandoma išspręsti GAN nestabilumą ir sustabdyti generatyvinio modelio persimokymą (angl. overfitting) duodant jam naujų tikslą. Vietoje tiesioginio diskriminatoriaus išvesties tobulinimo, generatorius bando sukurti duomenis, kurie atitinka tam tikras realių duomenų statistikas ar ypatybes, o diskriminatorius naudojamas tiktais tam, kad pateikti statistikas, kurias atrodo aktualu prilyginti.
2. Mažų partijų (angl. minibatch) diskriminacija – viena dažniausiai pasitaikančių generatoriaus problemų yra generatyvinio modelio užstrigimas prie vieno taško, kuriame jis sukuria vis tą pačią išvestį. Tai atsitinka, nes diskriminuojantis modelis vertina kiekvieną duomenų elementą atskirai ir neduoda generatoriui jokios priežasties didinti išvesties įvairumo. T. Salimans ir kt. siūlo tai spręsti pritaikant mažų partijų diskriminaciją – diskriminatoriui yra

duodamos mažos partijos duomenų egzempliorių, kurie yra vertinami kartu. Modelio išvestis vis tiek privalo išlikti viena skaitinė reikšmė, tad vertinimo metu yra klasifikuojamas vienas duomuo, tačiau kiti duomenys yra naudojami kaip pagalbinė informacija.

3. Istorinis vidurkis – prie GAN modelio netekties funkcijos funkcijos atitinkamai pridedami praeituose žingsniuose naudotų parametru θ_g ir θ_d vidurkiai. Taip bandoma išvengti gradiento orbitų (ciklų) aplink optimalų tašką.
4. Vienpusis žymų išlyginimas (angl. one-sided label smoothing) – naudojant šią metodiką, standartiniai klasifikavimo taikiniai, tikimybės 0 arba 1, yra pakeičiami labiau išlygintomis reikšmėmis kaip 0,1 ir 0,9. Taip yra padidinamas diskriminatoriaus atsparumas generatoriaus apgaudinėjimo bandymams ir padeda išvengti generatoriaus dominavimo.
5. Virtualus partijų normalizavimas – partijų normalizacija (angl. batch normalization) yra prietaikoma daugelyje giliojo mokymosi įgyvendinimų, mokymo stabilizavimui ir pagreitinimui. Jos metu modelio išvestys mokantis su atskiromis duomenų partijomis yra standartizuojamos taip, kad jų vidurkis būtų lygus 0, o standartinis nuokrypis 1. Tačiau tai reiškia, kad išvestys pagal vieną duomenų egzempliorių tampa priklausomas nuo kitų egzempliorių. Kad to išvengti, bet vis tiek išnaudoti normalizacijos naudą, autoriai pasiūlė virtualų partijų normalizavimą. Šiuo metodu kiekviena ižvestis yra normalizuojama pagal pavyzdinę partiją duomenų, kuri yra nustatoma prieš pradedant mokymo procesą.

Sekančią partiją patobulinimų 2018 metais pasiūlė T.Kerras ir kt. [KAL⁺18]. Nors jų parašytame straipsnyje siūlomi keli metodai modelių konkurencijos pagerinimui, pagrindinis pasiekimas buvo progresyvaus GAN auginimo metodas. Pagal jį, abu modeliai pradžioje yra mokomi su mažesnės rezoliucijos duomenimis, o mokymo eigoje ši rezoliucija yra progresyviai didinama. Tai ne tik sutaupo laiko (autoriai rašo, kad tai 2-6 kartus pagreitino modelių mokymą), bet ir padidina GAN stabilumą ir leidžia pasiekti daug geresnius rezultatus.

Šis laipsniškas rezoliucijos didinimas leidžia modeliui pirmiausiai atrasti bendrąsias duomenų savybes ir tada pamažu perkelti savo dėmesį link smulkesnių detalių, o ne bandyti viską išmokti vienu metu. Generatyvinis ir diskriminuojantis modeliai yra auginami vienodai ir visi sluoksniai yra prieinami viso mokymosi metu. Kai nauji sluoksniai yra pridedami, jie įvedami pamažu – naujam, padvigubintos rezoliucijos sluoksniui, yra pridedamas svoris, išreiškiamas kaip rezoliucijos daugiklis nuo 0,5 iki 1, pagal kurį sluoksnio rezoliucija yra sumažinama perpus ir mokymo metu tiesiškai didinama iki pilnos. Taip yra bandoma išvengti staigū pokyčių mokymo procese, kurie galėtų neigiamai jį paveikti.

Taip pat tame pačiame straipsnyje buvo pasiūlytas objektyviai įvertinamas modelio kokybės rodiklis vadinamas supjaustytu Wasserstein atstumu arba SWD (angl. Sliced Wasserstein Distance). Šis rodiklis nurodo duomenų tarpusavio skirtumą, tad vertinant generatyvinius modelius tikslas yra gauti kuo mažesnę šio rodiklio reikšmę lyginant sugeneruotus duomenis ir realius duomenis. Tai leidžia daug lengviau įvertinti gautus generatyvinius modelius, kadangi nereikia pasikliauti žmogiškuoju vertinimu, kuris yra nepastovus ir sunkiai įgyvendinamas dideliu mastu.

Visi šie pasiūlymai privėdė prie labai greito GAN tinklų tobulėjimo. Akivaizdžiausias to pavyzdys buvo veidų nuotraukų generavimo srityje, ką galima matyti 4 paveikslėlyje. Tačiau tas pats progresas gali būti pritaikomas ir kitose GAN taikymuose.



4 pav. GAN tinklų tobulėjimas veidų vaizdų kūrimo srityje [BAC⁺18]

2.3. 3D GAN

Trimačių modelių skaitmeninimas, pritaikant GAN algoritmus fotogrametrijos mokslui, nėra nagrinėjamas pirmą kartą. Keli metodai buvo iškelti kasdienių objektų, ar kultūrinio paveldo skaitmeninimui, turint ribotą kiekį informacijos. Darbo įkvėimas ir pagrindiniai literatūros apžvalgos šaltiniai buvo 2016 metais JiaJun Wu ir kt. parašytas straipsnis apie objektų trimatį skaitmeninimą iš vienos dvimatės nuotraukos [WZX⁺16] bei 2019 metais V.V. Kniaz ir kt. parašytas straipsnis apie bendresnį GAN pritaikymą architektūrinio paveldo srityje [KRK19].

2.3.1. IKEA baldų skaitmeninimas

JiaJun Wu ir kt. straipsnyje aprašomas trimačių objektų generavimas pritaikant tūrinės konvoluciujos neuroninius tinklus (angl. volumetric convolution) ir GAN. Pagrindinis autorių tikslas buvo sukurti metodą generuojantį 3D modelius, kurie yra panašūs į jau turimus realių objektų trimaciūs modelius tuo pačiu principu, kaip ir anksčiau minėti metodai kūrė naujus vaizdus pagal realias nuotraukas. Tačiau metadas taip pat buvo pritaikytas ir modelių generavimui pagal dvimatės nuotraukas. Tai padaryta ji sujungiant su variaciniu automatiniu koduotoju arba VAE (angl. Variational AutoEncoder) [KW13], [LSL⁺15].

Siūlomas 3D modelių generatorius paima 200 matmenų triukšmo vektorių, pripildytą atsitiktinėmis reikšmėmis iš intervalo $[0, 1]$ ir atvaizduoja jį į $64 \times 64 \times 64$ kubą. Tam naudojamas konvoluciūinis neuroninis tinklas su trimis $4 \times 4 \times 4$ konvoluciūiniai sluoksniai tiek generatyviniai, tiek diskriminuojančiame tinkle.

Kuriant modelius iš dvimačių nuotraukų pakeičiamas tiktais pradinis triukšmo vektorius. Viejoje atsitiktinių reikšmių yra naudojamos reikšmės aprašančios nuotraukos savybes. Šios reikšmės gaunamos naudojant variacinių automatinį koduotoją (VAE) – tai dirbtinis neuroninis tinklas, kuris paverčia jam duotą įvesties kintamajį į jį aprašantį daugiamtį statistinį reikšmių išdėstytmą (vektorius).

Straipsnyje naudojamas VAE tinklas sudarytas iš penkių erdvinių konvoluciinių sluoksnių, kurių dydžiai {11, 5, 5, 5, 8}, o žingsnių dydžiai (angl. Stride), nustatantys kaip algoritmas keliauja per paduodamus vaizdus, atitinkamai – {4, 2, 2, 2, 1}. Neskaitant partijų normalizacijos ir ReLU sluoksnių, kurie pritaikomi visuose metodo neuroniuose tinkle.

Netekties funkcija šiam metodui susideda iš trijų dalių: rekonstrukcijos netektis, t.y. skirtumo tarp sugeneruoto 3D modelio ir žinomo teisingo modelio – L_{recon} , jau minėtos GAN netekties – $L_{\text{3D-GAN}}$, ir Kullback-Leiber nuokrypio (angl. KL divergence) – L_{KL} . KL nuokrypis – tai skirtumas tarp dviejų tikimybų pasiskirtymų, šiuo atveju bandoma nenutolti nuo variacinio pasiskirstymo gauto su pradiniu vektoriumi $p(z)$ taip siekiant gauti modelį, kuris yra kuo panašesnis į nuotraukoje vaizduojamą objektą. Funkcija matoma formulėje (4) (α_1 ir α_2 KL nuokrypio ir rekonstrukcijos netekties svoriai):

$$L = L_{\text{recon}} + \alpha_1 L_{\text{KL}} + \alpha_2 L_{\text{recon}} \quad (4)$$

Metodo įvertinimui buvo naudojamas IKEA baldų trimačių modelių rinkinys, kuris yra gana sudėtingas, kadangi nuotraukos yra paimtos natūralioje erdvėje, todėl dažnai dalis objekto būna uždengta kitų objektų [LPT13]. Pritaikius GAN tinkle, gautas iki 13 % vidutinis patobulėjimas anksčiau geriausiems gautiems rezultatams (žr. 2 lentelę ir 5 pav.).

2 lentelė. Wu ir kt. straipsnyje [WZX⁺16] gauti 3D modeliavimo rezultatai (Tikslumas procentais).

Metodas	Lova	Knygų lentyna	Kėdė	Darbo stalas	Sofa	Stalas	Vidurkis
AlexNet-fc8 [GFR ⁺ 16]	29,5	17,3	20,4	19,7	38,8	16,0	23,6
AlexNet-conv4 [GFR ⁺ 16]	38,2	26,6	31,4	26,6	69,3	19,1	35,2
T-L Network [GFR ⁺ 16]	56,3	30,2	32,9	25,8	71,7	23,3	40,0
3D-VAE-GAN (geriausias rezultatas)	63,2	46,3	47,2	40,7	79,8	42,3	53,1



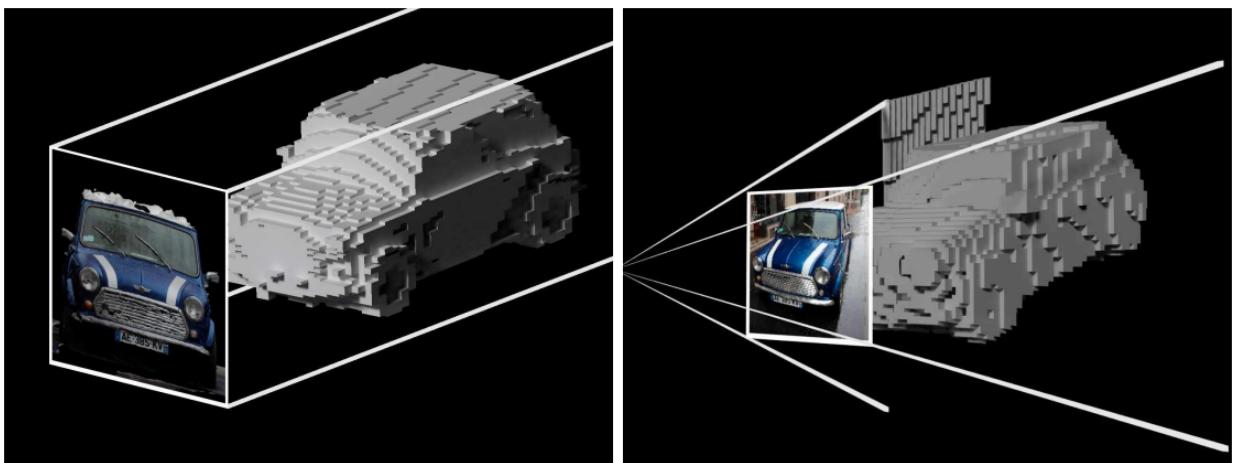
5 pav. Wu ir kt. straipsnyje [WZX⁺16] gauti 3D modeliavimo rezultatai.

2.3.2. Archeologinis paveldas

V. V. Kniaz ir kt. straipsnyje aprašomas GAN skaitmeninimo pritaikymas kultūrinio paveldo scenų rekonstrukcijai. Autoriai išsikelia problemą, kad nors modernūs tradicinės rekonstrukcijos metodai pasiekia įspūdingus rezultatus, dažnai yra turimos (išlikusios) tiktais viena ar kelios retos scenos ar objekto nuotraukos, tad yra reikalingas metodas, kuris galėtų gauti bent apytikslius 3D modelius su tokia ribota informacija.

Tam naudojamas Z-GAN tinklas [KKR18] ir sukuriamas naujas duomenų rinkinys sudarytas iš Neptūno šventyklos (Pestas, Italija), Cerere šventyklos (Pestas, Italija) ir Bostra archeologinės zonos (Sirija) 3D modelių ir nuotraukų.

Z-GAN metodas paremtas vokselinio modelio pjūviais. Vokselis – tai tūrinis pikselis, atvaizduojantis informaciją tam tikroje trimatės erdvės koordinatėje. Vokselinio modelio pjūvis – tai vienintelio gylio 3D modelio dalis gauta supjausčius modelį pagal vieną ašį. Autorių hipotezė buvo, kad atsižvelgus į pjūvių sutapimus su tam tikromis fotografuojamo objekto silueto dalimis galima patobulinti 3D skaitmeninimo GAN tinklais rezultatus. Tam reikia sulygiuoti turimas nuotraukas su viena iš mokymosi 3D modelių ašių. Kad šis lygiavimas būtų tikslus, fotografuojama erdvė yra įsivaizduojama kaip nupjauta piramidė (angl. frustum), kuri labiau atitinka kameros veikimą (žr. 6 pav.). Jos sujungimą su vokseliniu modeliu autorai vadina frukseliniu modeliu (angl. fruxel model).

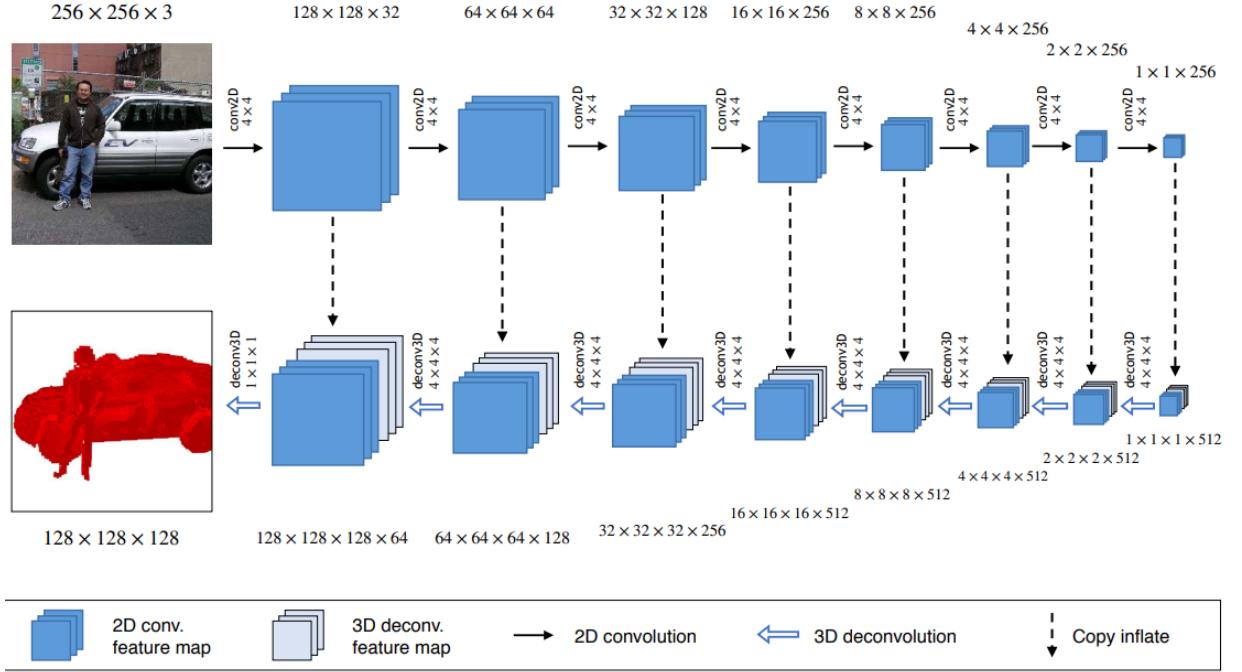


6 pav. Palyginimas tarp vokselinio (kairėje) ir frukselinio (dešinėje) modelio [KKR18].

Pagrindinė frukselinės modelio reprezentacijos mintis yra sutapatinti nuotraukose turimus kontūrus su modelio pjūviais ir taip generatorių išmokyti kurti modelius kurie labiau atitinka fotografuojamus objektus. Tokį modelį galima aprašyti naudojant keturių parametru aibę: $\{z_n, z_f, d, \alpha\}$, kur z_n yra kažkokia pradinė (artima) plokštuma, z_f galinė (tolima) plokštuma (z_f ir z_n skirtumas nusako modelio gylį), d – pjūvių skaičius, o α kameros matymo laukas (angl. field of view). Tokį frukselinį modelį galima konvertuoti į vokselinį padauginant kiekvieno pjūvio mastelių iš koeficiente k, matomo formulėje (5).

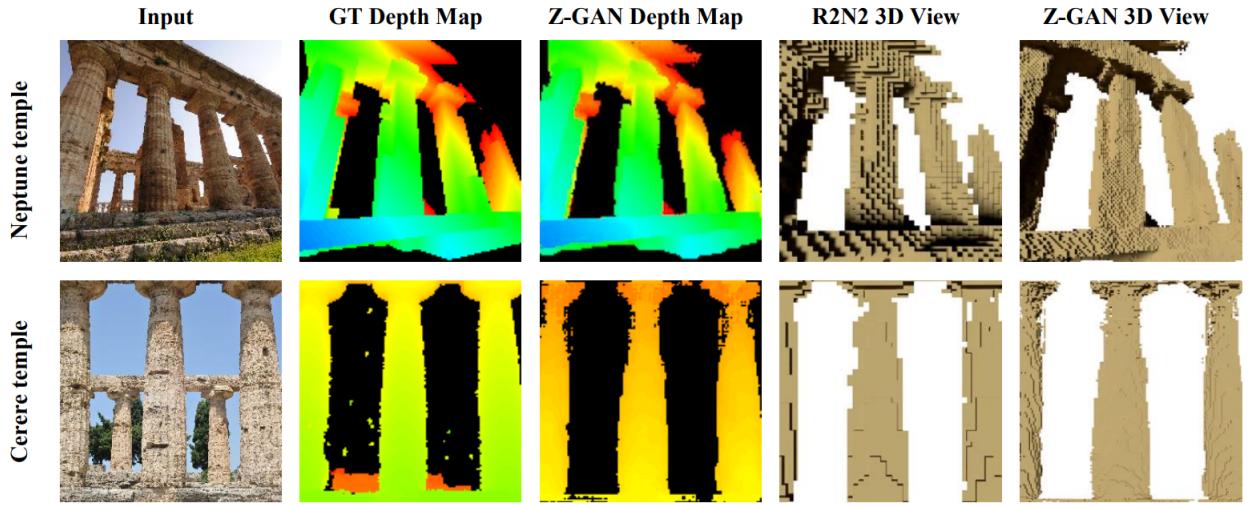
$$k = \frac{z_n}{z_n + s_z}; s_z = \frac{z_f - z_n}{d} \quad (5)$$

Z-GAN generatorius paima U-Net [RFB15] kaip pagrindą, kuris originalai buvo sukurtas nuotraukų segmentavimui. Autoriai pakeičia dvimačius dekonvoluciujos sluoksnius į trimačius ir pakoreguoja peršokimo sujungimus (angl. skip connections) taip, kad išlaikyti suderinamumą tarp 2D konvoluciujos ir 3D dekonvoluciujos sluoksniu. Generatoriaus architektūrą galima pamatyti 7 paveikslėlyje. Tuo tarpu diskriminuojantis modelis stebi tiktais atskirais pjūvius, o ne visą modelį, tad generatorius yra skatinamas juos teisingai atkurti.

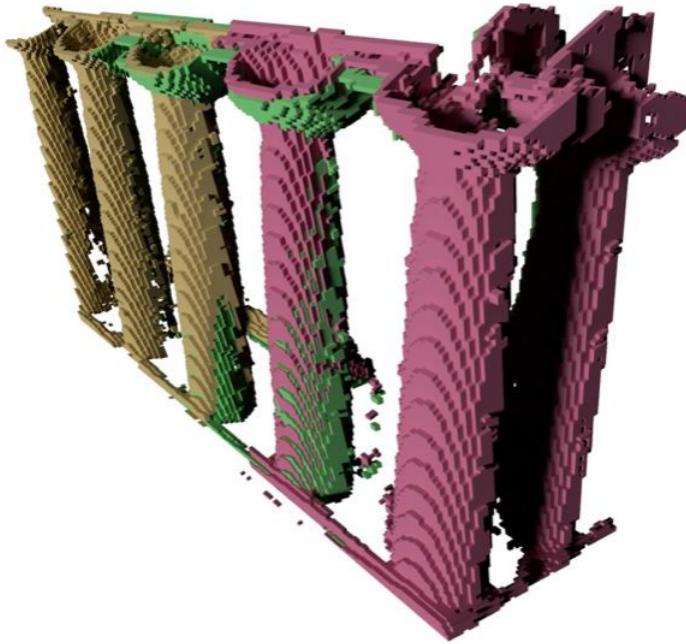


7 pav. Z-GAN generatyvinio modelio architektūra. [KKR18]

Pritaikius metodą kultūrinio ir architektūrinio paveldo nuotraukose esančių objektų skaitmeninimui (žr. 8 pav.) gauti aukštesnės raiškos rezultatai nei kituose darbuose ($128 \times 128 \times 128$ vietoje $64 \times 64 \times 64$), buvo galima atkurti mokymo metu nematytus objektus (pvz. arkas kurių nebuvvo mokymo duomenų rinkinyje) bei buvo įgyvendinta galimybė panaudoti kelias nuotraukas vieno trimacijo modelio kūrimui, taip gaunant geresnius rezultatus (žr. 9 pav.).



8 pav. Kniaz ir kt. straipsnyje [KRK19] gauti 3D modeliavimo rezultatai.



9 pav. Kniaz ir kt. straipsnyje [KRK19] gautas Cerere šventyklos 3D modelis sugeneruotas iš trijų gretimų nuotraukų.

Kelių nuotraukų panaudojimas yra pasiekiamas sujungiant kelis su viena nuotrauka gautos frukselinius modelius naudojant ICP algoritmą [CSS⁺02], kuris padeda suvaldyti skirtumus tarp kameros išorinių parametrų ir nuotraukų persiliejimą (angl. overlap).

2.4. 3D modelių kokybės tobulinimas

Vienas didžiausių 3D-VAE-GAN ir Z-GAN trūkumų yra jų gautų rezultatų vizualus nepatraukumas. Gauti 3D modeliai yra labai kampuoti ir mažos raiškos (rezoliucijos). Vienas to sprendimo būdas būtų dirbtinis raiškos padidinimas (angl. upscaling), tačiau 3D modelių raiškos didinimo metodai dar nėra plačiai ištyrinėti, daugiausiai pastangų yra kreipiama į 2D vaizdo raiškos didinimą, kaip NVIDIA DLSS (angl. Deep Learning Super Sampling), o esamos žinios ar straipsniai dažnai nėra viešai prieinami.

Iš karto generuojant aukštos raiškos modelius atsiranda atminties apribojimai, kadangi vokselių skaičius modeliuose auga kubine progresija. Li Sun ir kt. 2020 metais parašytame straipsnyje [SCX⁺20] siūlomas labiau atminties atžvilgiu efektyvesnis metodas HA-GAN, tačiau ir šis apsiriboja tiktais iki $256 \times 256 \times 256$ vokselių raiškos.

Kitas būdas padidinti šių modelių vizualų patrauklumą yra išlyginimas (angl. smoothing), kurio metu 3D modelių nelygumai galėtų būti ištiesinti ir perteiklinė informacija panaikinta. Vizualios kokybės tobulinimas gali sumažinti modelio atitikimą konkretiems fotografiuotiemis objektams, tad turėtų būti taikomas į tai atsižvelgiant, ypač tinkama kai vertinamas ne tikslus objekto atkūrimas, o objekto realistiškumas, kaip fono elementų kūrimas filmuose ar kompiuteriniuose žaidimuose.

2021 metų straipsnyje A. Bacciaglia ir kt. [BCL21] rašo apie paviršių išlyginimą 3D modelių topologijos optimizavimui. Topologijos optimizavimas yra atliekamas norint projektuoti lengvus ir efektyvius komponentus naudojamus pramoninės inžinerijos srityje, kaip lėktuvų ar automobilių

dalių kūrime. Autoriai siūlo metodą, kuris nepraranda tiek daug informacijos kiek kiti panašūs metodai, kurie dažnai užapvalina objektų kampus ir sukuria gamybai netinkamą išvestį. Tai gali būti pritaikoma ir 3D GAN metodams, kadangi dažniausiai bendra objekto struktūra yra sukuriama tinkamai, tad reikėtų kuo daugiau jos išlaikyti.

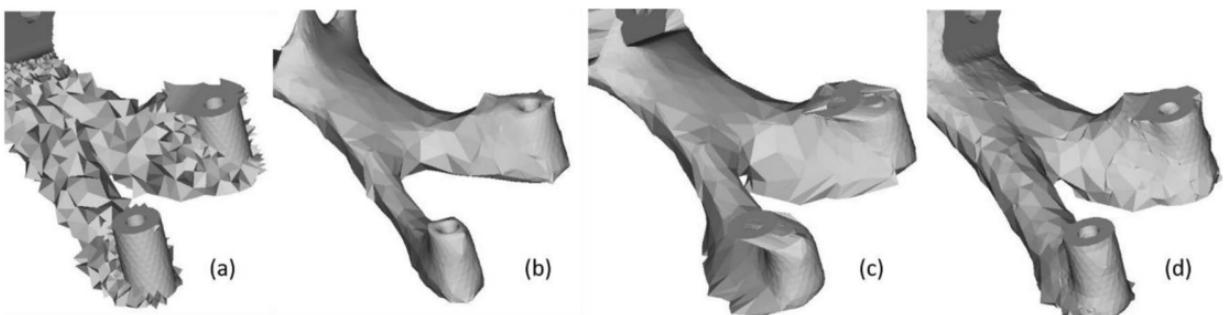
Autorių aprašomo metodo pagrindas yra Laplaso išlyginimas (angl. Laplacian smoothing), kuris per daugelį straipsnių ir iteracijų buvo optimizuojamas ir tobulinamas. Ši išlyginimo metodika keičia 3D tinklelio viršunių pozicijas pagal lokalią informaciją, pritaikant tokią Laplaso difuzijos lygtį:

$$\frac{\delta X}{\delta t} = \lambda L(X) \quad (6)$$

Formulėje (6) X yra viršunių rinkinys (angl. tensor), kuris bus pažingsniui keičiamas išlyginant 3D modelį. L – tai Laplaso funkcija, dažniausiai išreiškiama trimačiams duomenims kaip $\Delta_{\delta_x} i + \frac{\delta}{\delta y} j + \frac{\delta}{\delta z} k$, skaičiuojanti gradiento nuokrypi pagal koordinacijų x , y ir z dalines išvestines. λ – tai svoris nuo 0 iki 1 nusakantis difuzijos greitį, o δt nusako modelio paviršiaus tinklelio kitimą algoritmo eigoje.

Straipsnyje išvardinami keli patobulinimai paprastam Laplaso išlyginimui (gautų išlyginimo rezultatų palyginimą su kitu, tuo metu geriausiu laikomu, išlyginimo algoritmu galima matyti 10 paveikslėlyje):

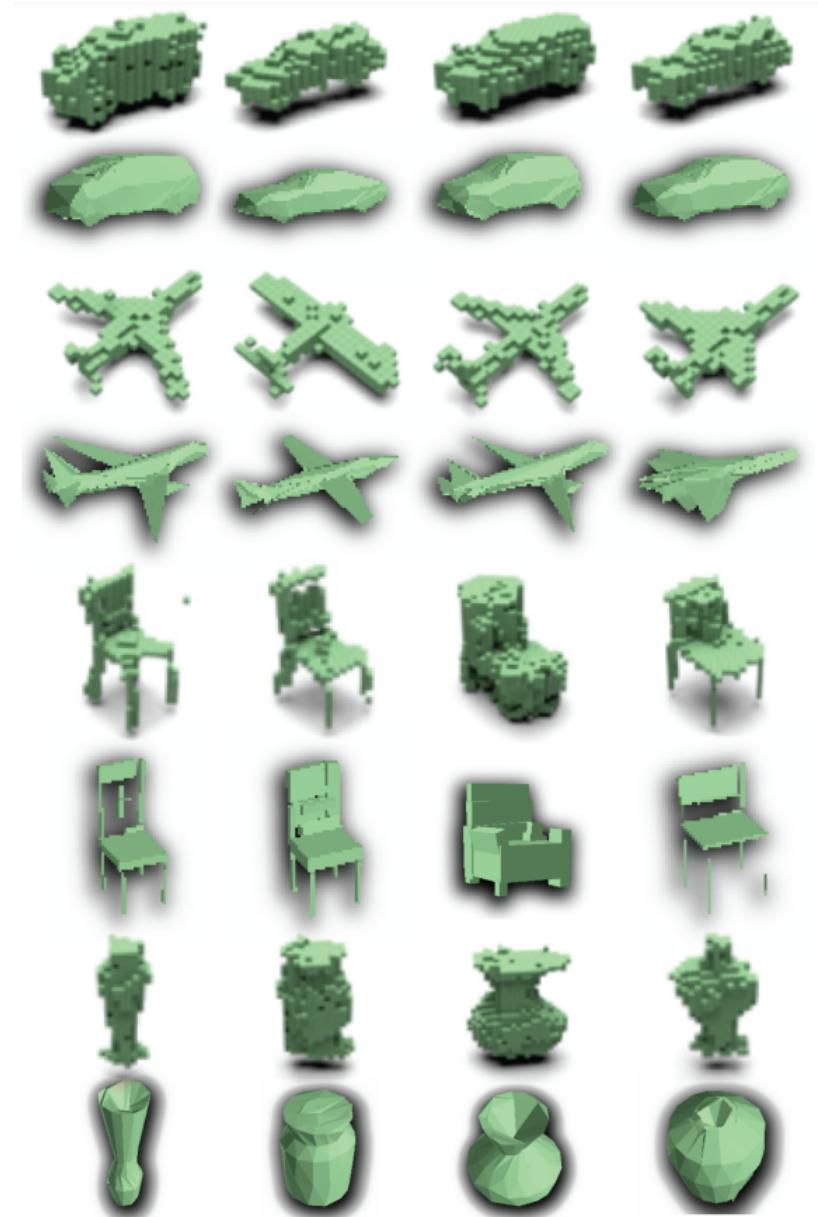
- Atsižvelgimas į mastelį, neleidžiant keisti tarp tinklelio susidariusių trikampių dydžio ir taip išvengiant per didelio viršunių išbarstymo.
- Atbulinio stūmimo pridėjimas, po kiekvienos iteracijos dalinai grąžinant viršunes link jų praėitos pozicijos pagal viršunės ir jos kaimynų vidutinį nuotoli.
- Lygių paviršių aptikimas – siekiama palikti jau esamas lygias (tiesias) modelio dalis nepaileistas, taip išvengiant nereikalingo kampų apvalinimo
- Skylių aptikimas – pritaikoma, kai modelyje yra tikimasi skilių ir jų savybių nereikia pakeisti
- Tūrio išlaikymas, kiekvienoje iteracijoje lyginant gautą modelio tūrį su pradiniu ir atitinkamai padauginant viršunių koordinates, kad išvengti objekto tūrio sumažėjimo.



10 pav. Bacciaglia ir kt. straipsnyje [BCL21] gauti išlyginimo rezultatai: (a) originalus modelis, (b) Taubino algoritmu gautas modelis [Tau99], (c) Taubino algoritmas neišlyginant tiesių paviršių, (d) Bacciaglia ir kt. algoritmas – optimizuotas HC-SDU

Šis algoritmas, kaip ir buvo minėta, yra pritaikomas turint trimatį taškų tinklelių (angl. 3D mesh), kurio tarpus užpildžius gaunamas tuščiaviduris 3D modelis. Tačiau su apžvelgtais 3D-GAN algoritmais yra gaunamas vokselinis modelis sudarytas iš daug mažų kubelių. Tad norint pritaikyti tokį išlyginimo algoritmą reikia gauti modelius atvaizduotus tinkleliu. Tai gali būti atliekama paprasčiausiai nurodant išorinius modelio vokselius kaip tinklelio viršunes arba galima pritaikyti 2019 metais P. Henderson ir V. Ferrari pasiūlytą 3D GAN algoritmą [HF19], kuris iš karto generuoja 3D tinkleliu atvaizduojamus modelius vietoje vokselinių modelių.

Šiame algoritme taip pat kuriami 3D modeliai iš vienos 2D nuotraukos, tačiau yra įvedami 3D tinklelio parametrai ir papildomai analizuojama fotografuotų objekto šešelių bei atspalvių informacija (angl. shading), kuri padeda geriau suprasti daiktų išlinkimus. Kaip matoma 11 paveikslėlyje, vien pritaikius šią metodą jau gaunami daug patrauklesni rezultatai, ypač išlenktiems objektams, tačiau juos galima toliau tobulinti pritaikant išlyginimą.



11 pav. Nelyginėse eilutėse – vokseliais grįstu 3D GAN algoritmu gauti rezultatai [GWM18]. Lyginėse eilutėse – tinkleliu grįsto 3D GAN rezultatai [HF19]

3. Išvada

Trimatis skaitmeninimas yra plačiai pritaikoma technologija, nuo skaitmeninių pramogų srities iki kultūrinio paveldo išsaugojimo. Šioje literatūros analizėje buvo apžvelgta dalis 3D skaitmeninimo metodų, koncentruojantis į GAN tinklų taikymą, bet peržiūrint ir tradicinius metodus. Iš surinktos informacijos toliau bus siekiamas magistrinio baigamojo darbo tikslas – įgyvendinti trimačių modelių išgavimo iš dvimačių nuotraukų metodą, naudojant besivaržančius neuroninius tinklus, kuris būtų pranašesnis lyginant su esamais metodais.

Šiuo metu geriausi 3D GAN skaitmeninimo iš vienos nuotraukos rezultatai siekia tiktai 128^3 ar 256^3 raišką (rezoliuciją) ir yra vizualiai nepatrauklūs, kampuoti. Tad aktualiausia kryptis pranašesnio metodo kūrimui – galutinio rezultato tobulinimas pritaikant papildomo apdorojimo (angl. post-processing) metodus kaip raiškos didinimas ar modelių išlyginimas.

Papildomai verta mėginti pritaikyti tradiciniuose metoduose naudojamas nuotraukų ypatybių išgavimo metodikas arba kitaip adaptuoti jų teoriją GAN mokymo eigoje. Taip pat, galima bandyti pritaikyti GAN mokymo patobulinimus, susilaikusius daug sėkmės 2D vaizdų srityje, kaip progresyvus auginimas, mažų partijų diskriminacija ar virtualus partijų normalizavimas.

Trimačių modelių generavimui iš kelių nuotraukų galima išmèginti kitus taškinį debesų apdorojimo metodus, nei Kniaz ir kt. straipsnyje [KRK19] naudotą ICP algoritmą. Be to, šiame straipsnyje buvo rodomas modelio kūrimas tiktai pagal viena kitai gretimas nuotraukas. Būtų pravartų turėti (sukurti) galimybę generuoti modelį pagal kelias nuotraukas padarytas iš skirtingų objekto pusų, taip suteikiant trūkstamą informaciją apie uždengtus elementus, kai turimos tiktai viena ar kelios nuotraukos darytos iš tos pačios pusės.

Literatūra

- [AOV12] Alexandre Alahi, Raphael Ortiz ir Pierre Vandergheynst. FREAK: Fast Retina Keypoint. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, p.p. 510–517, 2012. doi: 10.1109/CVPR.2012.6247715. URL: https://www.researchgate.net/publication/258848394_FREAK_Fast_retina_keypoint.
- [BAC⁺18] Miles Brundage, Shahar Avin, Jack Clark, Helen Toner ir k.t. The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. *ArXiv*, abs/1802.07228, 2018. URL: <https://arxiv.org/abs/1802.07228>.
- [BCL21] Antonio Bacciaglia, Alessandro Ceruti ir Alfredo Liverani. Surface smoothing for topological optimized 3D models. *Structural and Multidisciplinary Optimization*, 64, 2021-12. doi: 10.1007/s00158-021-03027-6. URL: https://www.researchgate.net/publication/353966431_Surface_smoothing_for_topological_optimized_3D_models.
- [BET⁺08] Herbert Bay, Andreas Ess, Tinne Tuytelaars ir Luc Van Gool. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359, 2008. issn: 1077-3142. doi: <https://doi.org/10.1016/j.cviu.2007.09.014>. URL: <https://people.ee.ethz.ch/~surf/eccv06.pdf>. Similarity Matching in Computer Vision and Multimedia.
- [CSS⁺02] D. Chetverikov, D. Svirko, D. Stepanov ir P. Krsek. The Trimmed Iterative Closest Point algorithm. *2002 International Conference on Pattern Recognition*, tom. 3, 545–548 vol.3, 2002. doi: 10.1109/ICPR.2002.1047997. URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.15.2020&rep=rep1&type=pdf>.
- [FB81] Martin A. Fischler ir Robert C. Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Commun. ACM*, 24(6):381–395, 1981. issn: 0001-0782. doi: 10.1145/358669.358692. URL: <https://doi.org/10.1145/358669.358692>.
- [FP10] Yasutaka Furukawa ir Jean Ponce. Accurate, Dense, and Robust Multiview Stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8):1362–1376, 2010. doi: 10.1109/TPAMI.2009.161. URL: https://www.researchgate.net/publication/44683582_Accurate_Dense_and_Robust_Multiview_Stereopsis.
- [GFR⁺16] Rohit Girdhar, David F. Fouhey, Mikel D. Rodriguez ir Abhinav Kumar Gupta. Learning a Predictable and Generative Vector Representation for Objects. *ArXiv*, abs/1603.08637, 2016. URL: <https://arxiv.org/abs/1603.08637>.

- [GPM⁺14] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville ir Yoshua Bengio. Generative Adversarial Nets. Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence ir K. Q. Weinberger, redaktoriai, *Advances in Neural Information Processing Systems*, tom. 27. Curran Associates, Inc., 2014. URL: <https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf>.
- [GWM18] Matheus Gadelha, Rui Wang ir Subhransu Maji. Multiresolution Tree Networks for 3D Point Cloud Processing, 2018. arXiv: 1807.03520. URL: <https://arxiv.org/abs/1807.03520>.
- [Har97] R.I. Hartley. In defense of the eight-point algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(6):580–593, 1997. doi: 10.1109/34.601246. URL: http://www.cs.cmu.edu/afs/scs/cs/15-463/f07/proj_final/www/amichals/fundamental.pdf.
- [HF19] Paul Henderson ir Vittorio Ferrari. Learning single-image 3D reconstruction by generative modelling of shape, pose and shading. *CoRR*, abs/1901.06447, 2019. arXiv: 1901.06447. URL: <http://arxiv.org/abs/1901.06447>.
- [Hir05] H. Hirschmuller. Accurate and efficient stereo processing by semi-global matching and mutual information. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, tom. 2, 807–814 vol. 2, 2005. doi: 10.1109/CVPR.2005.56. URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.88.8897&rep=rep1&type=pdf>.
- [Hop08] Hugues Hoppe. Poisson Surface Reconstruction and Its Applications. *Proceedings of the 2008 ACM Symposium on Solid and Physical Modeling*. Association for Computing Machinery, 2008. ISBN: 9781605581064. doi: 10.1145/1364901.1364904. URL: <https://doi.org/10.1145/1364901.1364904>.
- [HS88] C Harris ir M Stephens. A combined corner and edge detector. proc. of the 4th Alvey Vis. Conf, 1988. URL: <http://www.bmva.org/bmvc/1988/avc-88-023.pdf>.
- [YCH⁺13] Ming-Der Yang, Chih-Fan Chao, Kai-Siang Huang, Liang-You Lu ir Yi-Ping Chen. Image-based 3D scene reconstruction and exploration in augmented reality. *Automation in Construction*, 33:48–60, 2013. ISSN: 0926-5805. doi: <https://doi.org/10.1016/j.autcon.2012.09.017>. URL: <https://www.sciencedirect.com/science/article/pii/S0926580512001690>. Augmented Reality in Architecture, Engineering, and Construction.
- [IZZ⁺17] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou ir Alexei A. Efros. Image-to-Image Translation with Conditional Adversarial Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, p.p. 5967–5976, 2017. doi: 10.1109/CVPR.2017.632. URL: <https://arxiv.org/abs/1611.07004>.

- [KAL⁺18] Tero Karras, Timo Aila, Samuli Laine ir Jaakko Lehtinen. Progressive Growing of GANs for Improved Quality, Stability, and Variation. *International Conference on Learning Representations*, 2018. URL: <https://openreview.net/forum?id=Hk99zCeAb>.
- [KKR18] Vladimir A. Knyaz, Vladimir V. Knyaz ir Fabio Remondino. Image-to-Voxel Model Translation with Conditional Adversarial Networks. *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 2018-09. URL: https://openaccess.thecvf.com/content_eccv_2018_workshops/w6/html/Knyaz_Image-to-Voxel_Model_Translation_with_Conditional_Adversarial_Networks_ECCVW_2018_paper.html.
- [KL17] Ali Khaloo ir David Lattanzi. Hierarchical Dense Structure-from-Motion Reconstructions for Infrastructure Condition Assessment. *Journal of Computing in Civil Engineering*, 31(1):04016047, 2017. doi: 10.1061/(ASCE)CP.1943-5487.0000616. URL: <https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29CP.1943-5487.0000616>.
- [KRK19] Vladimir Knyaz, Fabio Remondino ir V. Knyaz. GENERATIVE ADVERSARIAL NETWORKS FOR SINGLE PHOTO 3D RECONSTRUCTION. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W9:403–408, 2019-01. doi: 10.5194/isprs-archives-XLII-2-W9-403-2019. URL: https://www.researchgate.net/publication/330766571_GENERATIVE_ADVERSARIAL_NETWORKS_FOR_SINGLE_PHOTO_3D_RECONSTRUCTION.
- [KW13] Diederik P Kingma ir Max Welling. Auto-Encoding Variational Bayes, 2013. doi: 10.48550/ARXIV.1312.6114. URL: <https://arxiv.org/abs/1312.6114>.
- [LCS⁺16] Yu-Fei Liu, Soojin Cho, B. F. Spencer ir Jian-Sheng Fan. Concrete Crack Assessment Using Digital Image Processing and 3D Scene Reconstruction. *Journal of Computing in Civil Engineering*, 30(1):04014124, 2016. doi: 10.1061/(ASCE)CP.1943-5487.0000446. URL: <https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29CP.1943-5487.0000446>.
- [Low04] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60:91–110, 2004. issn: 1573-1405. doi: 10.1023/B:VISI.0000029664.99615.94. URL: <https://doi.org/10.1023/B:VISI.0000029664.99615.94>.
- [LPT13] Joseph J. Lim, Hamed Pirsiavash ir Antonio Torralba. Parsing IKEA Objects: Fine Pose Estimation. *2013 IEEE International Conference on Computer Vision*, p.p. 2992–2999, 2013. doi: 10.1109/ICCV.2013.372. URL: https://people.csail.mit.edu/torralba/publications/ikea_iccv2013.pdf.

- [LSL⁺15] Anders Boesen Lindbo Larsen, Søren Kaae Sønderby, Hugo Larochelle ir Ole Winther. Autoencoding beyond pixels using a learned similarity metric, 2015. doi: 10 . 48550/ARXIV.1512.09300. URL: <https://arxiv.org/abs/1512.09300>.
- [LTH⁺17] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew P. Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang ir Wenzhe Shi. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*:105–114, 2017. URL: <https://arxiv.org/abs/1609.04802>.
- [MY09] Jean-Michel Morel ir Guoshen Yu. ASIFT: A New Framework for Fully Affine Invariant Image Comparison. *SIAM Journal on Imaging Sciences*, 2(2):438–469, 2009. doi: 10 . 1137/080732730. URL: <https://doi.org/10.1137/080732730>.
- [ML18] Zhiliang Ma ir Shilong Liu. A review of 3D reconstruction techniques in civil engineering and their applications. *Advanced Engineering Informatics*, 37:163–174, 2018. ISSN: 1474-0346. doi: <https://doi.org/10.1016/j.aei.2018.05.005>. URL: <https://www.sciencedirect.com/science/article/pii/S1474034617304275>.
- [Nis04] D. Nister. An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6):756–770, 2004. doi: 10 . 1109/TPAMI . 2004 . 17. URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.86.8769&rep=rep1&type=pdf>.
- [PHL⁺17] Ashok Kumar Patil, Pavitra Holi, Sang Keun Lee ir Young Ho Chai. An adaptive approach for the reconstruction and modeling of as-built 3D pipelines from point clouds. *Automation in Construction*, 75:65–78, 2017. ISSN: 0926-5805. doi: <https://doi.org/10.1016/j.autcon.2016.12.002>. URL: <https://www.sciencedirect.com/science/article/pii/S0926580516304745>.
- [RD06] Edward Rosten ir Tom Drummond. Machine Learning for High-Speed Corner Detection. Aleš Leonardis, Horst Bischof ir Axel Pinz, redaktoriai, *Computer Vision – ECCV 2006*, p.p. 430–443, Berlin, Heidelberg. Springer Berlin Heidelberg, 2006.
- [RFB15] Olaf Ronneberger, Philipp Fischer ir Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. *CoRR*, abs/1505.04597, 2015. arXiv: 1505 . 04597. URL: <http://arxiv.org/abs/1505.04597>.
- [RGL⁺14] Pablo Rodriguez-Gonzalvez, Diego Gonzalez-Aguilera, Gemma Lopez-Jimenez ir Immaculada Picon-Cabrera. Image-based modeling of built environment from an unmanned aerial system. *Automation in Construction*, 48:44–52, 2014. ISSN: 0926-5805. doi: <https://doi.org/10.1016/j.autcon.2014.08.010>. URL: <https://www.sciencedirect.com/science/article/pii/S0926580514001897>.

- [SBM⁺15] Jūratė Sužiedelytė-Visockienė, Renata Bagdžiūnaitė, Naglis Malys ir Vida Maliene. Close-range photogrammetry enables documentation of environment-induced deformation of architectural heritage. *Environmental Engineering and Management Journal*, 14, 6, 2015. URL: <http://eprints.nottingham.ac.uk/id/eprint/34653>. Full text of article not available online. No doi. Permission given by journal publishers to deposit in repository. KJB 11.07.2016.
- [SC13] Changhun Sung ir Myung Jin Chung. Multi-Scale Descriptor for Robust and Fast Camera Motion Estimation. *IEEE Signal Processing Letters*, 20(7):725–728, 2013. doi: 10.1109/LSP.2013.2264672.
- [SCX⁺20] Li Sun, Junxiang Chen, Yanwu Xu, Mingming Gong, Ke Yu ir Kayhan Batmanghelich. Hierarchical Amortized Training for Memory-efficient High Resolution 3D GAN, 2020. doi: 10.48550/ARXIV.2008.01910. URL: <https://arxiv.org/abs/2008.01910>.
- [SGZ⁺16] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, Xi Chen ir Xi Chen. Improved Techniques for Training GANs. D. Lee, M. Sugiyama, U. Luxburg, I. Guyon ir R. Garnett, redaktoriai, *Advances in Neural Information Processing Systems*, tom. 29. Curran Associates, Inc., 2016. URL: <https://proceedings.neurips.cc/paper/2016/file/8a3363abe792db2d8761d6403605aeb7-Paper.pdf>.
- [SKK15] Hyojoo Son, Changmin Kim ir Changwan Kim. Fully Automated As-Built 3D Pipeline Extraction Method from Laser-Scanned Data Based on Curvature Computation. *Journal of Computing in Civil Engineering*, 29(4):B4014003, 2015. doi: 10.1061/(ASCE)CP.1943-5487.0000401. URL: [https://ascelibrary.org/doi/abs/10.1061/\(ASCE\)CP.1943-5487.0000401](https://ascelibrary.org/doi/abs/10.1061/(ASCE)CP.1943-5487.0000401).
- [Tau99] Gabriel Taubin. A Signal Processing Approach To Fair Surface Design. *Computer Graphics (Proceedings of Siggraph '95)*, 29, 1999-07. doi: 10.1145/218380.218473. URL: https://www.researchgate.net/publication/2374153_A_Signal_Processing_Approach_To_Fair_Surface_Design.
- [TMH⁺00] Bill Triggs, Philip F. McLauchlan, Richard I. Hartley ir Andrew W. Fitzgibbon. Bundle Adjustment — A Modern Synthesis. *Vision Algorithms: Theory and Practice*, p.p. 298–372, 2000. ISBN: 978-3-540-44480-0.
- [WZX⁺16] Jiajun Wu, Chengkai Zhang, Tianfan Xue, Bill Freeman ir Joshua B. Tenenbaum. Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling. *NIPS*, 2016. URL: <https://arxiv.org/abs/1610.07584>.
- [ZHK⁺14] Stefanie Zollmann, Christof Hoppe, Stefan Kluckner, Christian Poglitsch, Horst Bischof ir Gerhard Reitmayr. Augmented Reality for Construction Site Monitoring and Documentation. *Proceedings of the IEEE*, 102:137–154, 2014-02. doi: 10.1109/JPROC.2013.2294314. URL: <https://www.researchgate.net/publication/2294314>.

261045453_Augmented_Reality_for_Construction_Site_Monitoring_and_Documentation.

- [ZXL⁺17] Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang and Dimitris N. Metaxas. StackGAN: Text to Photo-Realistic Image Synthesis with Stacked Generative Adversarial Networks. *2017 IEEE International Conference on Computer Vision (ICCV)*:5908–5916, 2017. URL: <https://arxiv.org/abs/1612.03242>.