LATVIJAS UNIVERSITĀTES
DATORIKAS FAKULTĀTE

STEREOSKOPISKA VIDEO MATERIĀLA ĢENERĀCIJA NO PARASTIEM VIDEO MATERIĀLIEM, IZMANTOJOT MAŠĪNMĀCĪŠANOS

BAKALAURA DARBS

Autors: **Kārlis Grošs**
Studenta apliecības Nr.: kg14021
Darba vadītājs: Dr. dat. Pēteris Paikens

RĪGA 2018

ANOTĀCIJA

Mūsdienās virtuālās realitātes tehnoloģijas kļūst aizvien populārākas. Tomēr, satura veidošanas sarežģītības dēļ kvalitatīvi trīs dimensiju jeb stereoskopiski video materiāli, ko izmantot virtuālajā realitātē, joprojām nav plaši pieejami. Darba autors pēta iespēju atvieglot stereoskopisku video materiālu izveidi, izmantojot mākslīgos neironu tīklus, kuri šādus materiālus ģenerētu no parasta divu dimensiju platleņķa video. Tiek apskatīti iepriekšējie mēģinājumi risināt līdzīgas problēmas, no kuriem viens tiek pārveidots, uztrenēts ar virtuālās realitātes materiāliem un empīriski izvērtēts. Tiek secināts, ka šāda metode ir īstenojama un praktiski pielietojama.

Atslēgvārdi: Virtuālā Realitāte, Attēlu apstrāde, Stereoskopija, Mākslīgie neironu tīkli

ABSTRACT

STEREOSCOPIC VIDEO GENERATION FROM GENERIC VIDEOS USING MACHINE LEARNING

As Virtual Reality (VR) technologies are gaining popularity, the demand for good quality stereoscopic VR video content is rising, but their production remains challenging. This paper explores the possibility to generate such content from regular wide-angle videos using artificial neural networks. Previous work on related problems is examined, and one solution is adapted and retrained for this purpose. It is empirically concluded that such content generation method is indeed feasible.

Keywords: Virtual Reality, Image processing, Stereoscopy, Artificial neural networks

SATURS

Apzīmējumu saraksts 6

Ievads 8

Virtuālā realitāte un VR video attēlošana 8

Mašīnmācīšanās 9

Pētāmā problēma 9

Darba mērķis 10

Hipotēzes 10

Saistītās problēmas 10

1. Telpiskas vides attēlošanas un uztveres pamatprincipi dabā un virtuālajā realitātē 12

1.1. Stereoskopija, disparitāte un fokuss 12

1.2. Perspektīva un projekcija 13

1.3. Papildinformācija, pieredze 13

1.4. Kompozīcija 14

1.5. Kustība 14

1.6. Virtuālās realitātes vides attēlošanas modelis 15

1.7. Virtuālai realitātei paredzētā satura īpašības 15

2. Attēlu apstrādes neironu tīklu pamatprincipi 17

2.1. Mākslīgais neironu tīkls 17

2.2. Konvolūciju tīkli 17

2.3. Attēla ievaddatu interpretācija neironu tīklā 18

2.4. Tīkla dziļums, iezīmju atpazīšana, apkopojošie slāņi 19

2.5. Izšķirtspējas saglabāšana, pārleciena slāņi, konkatenācija, transponētās un dziļumā atdalāmās konvolūcijas 20

2.6. Disparitātes, gradienta un dziļuma slāņi 21

2.7. Vērtības zuduma funkcija 22

2.8. Virtuālās realitātes satura attēlojums neironu tīklā 22

3. Risinājuma modelis 24
3.1. Modeļa sākotnējā struktūra 24
3.2. Ieejas datu pirmsapstrāde 25
3.3. Modeļa struktūras variācijas 25
3.4. Modeļa pētītā papildfunkcionalitāte 26
3.5. Tehniskie ierobežojumi 27
3.6. Implementācija 27
3.7. Modeļa pielietošana 27
4. Tīklu trenēšana 29
4.1. Izmantotie treniņdati 29
4.2. Datu augmentācija 29
4.3. Treniņdatu pirmsapstrāde 30
4.4. Trenēšanas parametri 30
4.5. Trenēšanas un validācijas datu sadalījums 31
4.6. Optimizātors, vērtības zuduma funkcija un kļūdu aprēķins 31
4.7. Trenēšanas ilgums 32
Rezultāti un diskusija 33
Modeļa un trenēšanas mainīgo ietekme uz rezultātu 33
Veiktspēja 36
Salīdzinājums ar citu autoru darbiem 36
Kvalitatīvs vērtējums 36
Labākais risinājums 37
Pozitīvi piemēri 37
Negatīvi piemēri 39
Secinājumi 40
Hipotēzes 40
Iespējamie uzlabojumi un tālākie pētījumi 41
Izmantotās literatūras saraksts 42

Pielikumi 46
1. Disparitāšu slāņu apvienojuma salīdzinājums dažādiem modeļiem 46
Dokumentārā lapa 48

APZĪMĒJUMU SARAKSTS

Autostereogramma - paņēmiens, kā aplūkot stereoskopisku attēlu vai video bez papildu palīglīdzekļiem, palielinot vai pamazinot ierasto leņķi starp acīm.

Datorredze - datoru analoģija cilvēka redzei, mākslīgā intelekta un mašīnmācīšanās novirziens.

Datu augmentācija - datu kopas palielināšana, pavairojot esošos datus, katram dublikātam izmanot kādu mainīgo.

Dropout - neironu tīkla slānim piešķirts daļskaitlis, kurš norāda, cik liela daļa nejauši izvēlētu neironu konkrētajā slāni trenēšanas laikā tiks atslēgti. Atslēgto neironu sadalījums mainās katrā trenēšanas solī. Nodrošina *overfitting* parādības samazināšanos.

Gradients - izmaiņmērs. Izsaka vērtības izmaiņu attiecībā pret blakus vērtībām.

Hiperparametrs - neironu tīkla parametrs, kuru nosaka tīkla veidotājs, nevis pats tīkls trenējoties

Iezīmju dimensija (mākslīgos neironu tīklos) - neironu tīkla slāņa sastāvdaļa, apzīmē neatkarīgu mainīgo vai to kopu slānī.

Interpolācija - vērtības iegūšana starp divām diskrētām vērtībām, zinot attālumu no tām.

Konvolūcija - vienas elementu secības pārveidošana otrā saskaņā ar kādu noteiktu algoritmu.

Konvolūciju kodols - konvolūciju darbības vērtību apgabals, nosaka elementus, no kuriem tiks aprēķināta vērtība.

Mašīnmācīšanās - datorikas nozare, kas pēta paņēmienus, kā datoriem likt adaptēties un risināt problēmas, iepriekš tām nedefinējot precīzu risinājuma algoritmu.

Mākslīgais neironu tīkls - vispopulārākais mašīnmācīšanās rīks. Konceptuāli līdzīgs bioloģiskajiem neironu tīkliem.

Mediji - informācijas nesēji, informatīvā vide

Mērogošana - kāda priekšmeta izmēra izmaiņa, nemainot tā sastāvu vai proporcijas.

Monoskopisks - kaut kas, kas aplūkots no viena skatu punkta, divās dimensijās.

Neirona aktivācija - iekš neironu tīkla notiekoša signāla padošana no viena neirona citā.

Optiski kropļojumi - projekcijas vai optikas novirze no ideāla, rektilineārās projekcijas.

Overfitting - parādība, ka trenēšanas laikā konkrētiem testpiemēriem tiks atrasta kāda nejauša sakarība, kas ļauj labāk paredzēt konkrētos testpiemērus, bet nepalīdz vai pat traucē vispārīgajā risinājumā

Pikselis - vismazākā rastra attēla vienība, visi rastra attēli tiek konstruēti no daudziem dažādu krāsu pikseļiem.

Projekcija - telpiska trīs dimensiju ķermeņa attēlojums divās dimensijās

Rektilineārā projekcija - ģeometriska projekcija, kur katra taisnā līnija dabā ir taisna arī projekcijā, neveido optiskus kropļojumus

Skalārs lielums - 0-dimensionāls lielums, tieši viens skaitlis.

Slānis - struktūra mākslīgajos neironu tīklos. Tīkls sastāv no vairākiem slāņiem, slāņi sastāv no neironiem.

Stereoskopija - metode, kā uztvert un attēlot trīs dimensiju telpu ar diviem divdimensiju attēliem.

Šūpes kaite - vestibulārā aparāta problēma, ko izraisa redzētā neatbilstība justajām kustībām.

Trenēšanas solis - vismazākais cikls neironu tīklu trenēšanā, katrā solī caur tīklu tiek izlaists noteikts skaits datu elementu,

Validācija - korekcijas pārbaude, neironu tīklos - tīkla paredzēšanas spēju pārbaude, metrika, pēc kuras izvēlas, vai izmaiņa bija lietderīga

VR - virtuālā realitāte, datorā veidota mākslīga trīsdimensiju vide, kas rada realitātes ilūziju. Lai panāktu šo ilūziju, tiek izmantoti speciāli stereoskopiska efekta displeji (brilles) un palīgierīces manipulācijām ar virtuālās realitātes objektiem.

IEVADS

Pagājušā gadsimta sākumā parādījās un pakāpeniski ieguva popularitāti jauna multimediju forma - filmas. Tā kā tā spēja attēlot notikumus daudz precīzāk, informatīvāk un tuvāk realitātei, nekā teksts, skaņas ieraksti vai bildes, tādēļ ar laiku filmas kļuva par visizplatītāko izklaides industrijas mediju.

Gadsimta vidū tika izgudrota arī televīzija, kas nodrošina tās pašas priekšrocības pār citiem mediju veidiem kā filmas, bet ļāva videomateriālu pārraidīt attālināti, nodrošinot ātru un apjomīgu mediju pieejamību. Jau kopš pagājušā gadsimta astoņdesmitajiem gadiem televīzija ir visplašāk izmantotā no multimediju veidiem.

Līdz ar gadsimta miju popularizējās arī interneta mediji. Ja iepriekš videomateriālu pieejamība bija atkarīga no konkrēta grafika televīzijā vai fizisku priekšmetu - datu nesēju pieejamības filmām, tad internetā, ja publicētājs to vēlas, jebkurš var piekļūt jebkuram video materiālam jebkurā laikā, paliekot savās mājās.

Turklāt, ja kādreiz video materiālus pārsvarā veidoja tikai speciālisti, tad līdz ar kvalitatīvas filmēšanas funkcijas parādīšanos viedtālruņos un video dalīšanās platformu izveidošanos, video materiālus var uzfilmēt un publicēt gandrīz jebkurš.

Līdz ar to, mūsdienās gandrīz jebkuram cilvēkam ir pieejama neaptverami plaša un daudzveidīga video multimediju vide.

Virtuālā realitāte un VR video attēlošana

Pēdējos gados strauji attīstās un ir kļuvusi populāra jauna datorikas nozare - virtuālā realitāte (VR). Tajā pēta tehnoloģijas, ar kurām var radīt realitātes ilūziju mākslīgā vidē. [1, 2] Ja kādreiz šādas tehnoloģijas bija ļoti dārgas un reti demonstrēja pārliecinošus rezultātus, tad mūsdienās ir pieejami kvalitatīvi virtuālās realitātes risinājumi, kuri ir pietiekami lēti un vienkārši izmantojami, lai tos varētu lietot lielākā daļa cilvēku.

Viens no virtuālās realitātes pielietojumiem ir video attēlošana. Virtuālās realitātes ierīcē attēlots, kvalitatīvi veidots video cilvēkam rada daudz spēcīgāku klātbūtnes sajūtu, nekā parasts, uz ekrāna attēlos video materiāls, un, līdz ar to, tas sniedz lielāku izklaides potenciālu un tajā redzamā informācija ir intuitīvāk uztverama.

Tomēr, lai virtuālās realitātes ierīcē attēlotais video materiāls izskatītos ticams, ir nepieciešama dziļuma informācija, kas nav sastopama parastos divu dimensiju video. Mūsdienās šo problēmu vislabāk var risināt ar stereoskopisku video formātu - katrai acij tiek

attēlots savs video fragments, kurš attēlo notiekošo no nedaudz atšķirīga skatu punkta, un, tos salīdzinot, smadzenes uztver trīs dimensiju vidi.

Diemžēl, ierīces, kas spēj filmēt attiecīgu stereoskopisku video materiālu, joprojām ir dārgas un nav plaši pieejamas [3], un pat ja tās varētu izmantot, tas nozīmētu, ka plašie agrāk veidotie video resursi nevar tikt kvalitatīvi attēloti virtuālās realitātes ierīcēs, kuras attēlo tikai stereoskopisku video materiālu.

Ar jau esošu, speciālu programmatūru ir iespējams manuāli izveidot stereoskopiskus video materiālus no parasta video [4], bet tas ir ļoti darbietilpīgs process, un iegūtais video visdrīzāk neatbildīs kvalitātes prasībām, kuras nepieciešamas, lai cilvēka smadzenes pareizi uztvertu telpiskumu virtuālās realitātes vidē.

Mašīnmācīšanās

Mūsdienās ļoti strauji attīstās arī cita datorikas nozare - mašīnmācīšanās. Atšķirībā no tradicionālās pieejas, kurā datori problēmu risināšanu veic ar iepriekš programmētu un stingri definētu algoritmu palīdzību, mašīnmācīšanās gadījumā algoritms netiek tieši definēts, bet datoram tiek iedots pietiekami liels piemēru apjoms, lai tas pats varētu atrast optimālo algoritmu, bez nepieciešamības programmētājam pirms tam pilnībā izprast un atrisināt doto problēmu.

Viens no ļoti izplatītiem un veiksmīgiem mašīnmācīšanās pielietojumiem ir attēlu un video apstrāde, jo tie satur lielu daudzumu sarežģīti strukturējamas informācijas, un atrast deterministisku algoritmu, kas strādā ar attēliem vispārīgā gadījumā ir ļoti sarežģīti vai pat praktiski neiespējami.

Darba autors uzskata, ka video materiāla apstrāde ar mašīnmācīšanos var sniegt risinājumu iepriekš minētajai problēmai par VR video pieejamību.

Pētāmā problēma

Ņemot vērā pašreizējo situāciju, darba autors izvirzīja mērķi izpētīt iespēju ģenerēt stereoskopisku video materiālu no parasta video materiāla, izmantojot mašīnmācīšanās, ar nolūku šo video materiālu pēc tam izmantot virtuālās realitātes vidē.

Darbā tiek aplūkota metode, kā to izdarīt, tiek pētīts, kādā kvalitātē ar to var iegūt attiecīgi uzģenerēto stereoskopisko materiālu, kā šo metodi var uzlabot un kādas problēmas jārisina, lai to izdarītu. Uzlabojumi tiek empīriski pārbaudīti.

Darba mērķis

Atrast risinājumu, ar kuru var uzģenerēt stereoskopisku attēlu no monoskopiska platleņķa attēla. Izpētīt dažādas šī risinājuma variācijas, empīriski tās novērtēt. Secināt, vai šāds risinājums ir potenciāli pielietojams VR video ģenerācijai, ja nav, tad jānoskaidro kādēļ, un kas ir jāmaina.

Hipotēzes

Tika izvirzītas sekojošas hipotēzes:

- Ar šo metodi ir iespējams iegūt cilvēkam stereoskopiski interpretējamu otras acs attēlu;
- Šo metodi var veiksmīgi pielietot VR video veidošanā;
- Atmiņas un trenēšanas laika ierobežojumi neļauj šo tieši izdarīt praktiskā 1920x1080 izšķirtspēja;
- Ģenerējot VR video materiālus rezultējošā kadra vidējā absolūtā kļūda būs vismaz divas reizes lielāka, nekā ģenerējot parastus stereoskopiskus kadrus ne-VR 3D filmām.

Saistītās problēmas

Pētot un risinot doto problēmu, tiek apskatītas parādības un pētītas metodes, kuras ir nozīmīgas arī citām problēmām.

Ja katra kadra ģenerēšanai video materiālā tiek izmantots tikai viens kadrs, tad to pašu metodi var izmantot arī, lai no parastiem fotoattēliem ģenerētu stereoskopiskas bildes.

Ja no video materiāla ir iespējams uzkonstruēt kvalitatīvu attēlojumu tam pašam saturam no cita skatu punkta, tad tas nozīmē, kad ir iegūta arī dziļuma un trīs dimensiju telpas informācija par dotā video saturu. Šo informāciju varētu izmantot, lai izveidotu trīs dimensiju modeļus objektiem, kas redzami dotajā materiālā.

Tā kā telpiska informācija palīdz noteikt kameras atrašanos telpā, un kvalitatīvai skatu punkta koriģēšanai arī ir nepieciešama dziļuma informācija, pētāmā problēma var saturēt būtiskas sastāvdaļas risinājumam, kas stabilizētu kameras kustību video materiālam, kas filmēts apstākļos, kuros neizdevās noturēt kameru vienā pozīcijā vai vienmērīgā kustībā, bet rezultējošais materiāls ir vēlams bez liekām kustībām.

Papildus jau minētajiem pielietojumiem, telpiska informācija un skatu punkta korekcijas iespēja var ļaut palielināt izšķirtspēju video fragmentam, kā papildinformāciju katram video kadram izmantojot tikai citus tā paša fragmenta kadrus. Tāpat var darīt arī pretējo - samazināt informācijas daudzumu, nemainot izšķirtspēju, jeb, veikt kompresiju.

1. TELPISKAS VIDES ATTĒLOŠANAS UN UZTVERES PAMATPRINCIPI DABĀ UN VIRTUĀLAJĀ REALITĀTĒ

Lai gan mēs visi dzīvojam telpiskā, trīs dimensiju pasaulē, neviens cilvēks tiešā veidā šīs trīs dimensijas nemaz neredz, jo cilvēka acs tīklene, gluži kā jebkurš fotosensors vai cita optiska ierīce, redz pasaules projekciju tikai divās dimensijās. Pašu trīs dimensiju telpas uztveri rada smadzenes, apstrādājot no acīm iegūto informāciju. Līdzīgus attēlu apstrādes paņēmienus izmanto arī datorredzē. Šajā nodaļā ir apkopoti izplatītākie no šiem paņēmieniem un pamatprincipi, uz kuriem tie balstās.

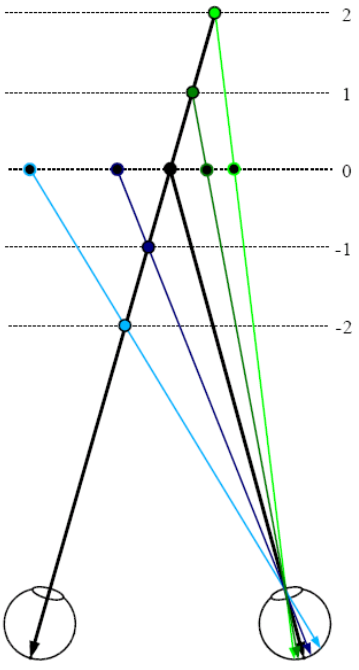
1.1. Stereoskopija, disparitāte un fokuss

Viens no visefektīvākajiem un viennozīmīgi bioloģiski visvairāk izmantotais veids, kā iegūt telpisku informāciju no attēliem ir stereoskopija. [5] Ar stereoskopiju telpisku informāciju iegūst salīdzinot divus attēlus. Jo lielāka ir kāda objekta nobīde starp abiem attēliem, jo tuvāk atrodas objekts. Ar triangulācijas palīdzību šo attālumu var arī aprēķināt. [6]

Datorgrafikā šādas nobīdes apzīmē ar disparitātes matricu. Katram attēla pikselim tiek atrasts attiecīgais pikselis otrā attēlā, un tiek aprēķināta nobīde attiecībā pret otru bildi (Skat. *1.1. att.*). Katrs elements, kurš atrodas šajā matricā, ar savu vērtību norāda, cik tālu pa kreisi vai pa labi kāds pikselis ir atrodams otrā attēlā. [7]

Šādu matricu izrēķināt ir netriviāli, jo katram pikselim ir iespējamas vairākas šādas nobīdes, un vajag ņemt vērā apkārtējo vidi, lai noskaidrotu, kura ir īstā. Tas disparitātes matricas deterministisku aprēķinu padara par ļoti algoritmiski sarežģītu un praktiski neizmatojamu. [8, 9]

Tomēr, cilvēka smadzenes šo funkciju pilda ļoti labi, pieņemami interpretējot pat neatbilstošus un nepilnīgus attēlus. Piemēram, arī cilvēks, kuram viena acs ir daudz tuvredzīgāka par otru, joprojām tālumā spēj noteikt dziļumu. [10]



1.1. att. *Dziļuma ietekme uz punkta nobīdi*

Vēl stereoskopiska attēla dziļuma datu noteikšanai ir svarīga arī fokusa plakne. Tā ir plakne, kas ir perpendikulāra skatiena virzienam, un uz kuras objekti cilvēkam vai ierīcei attēlosies visasāk. Fokusa plaknes vidū ir fokālais punkts, jeb fokuss. Uz to tiek konverģētas cilvēka acis. Līdz ar to, rēķinot precīzu disparitātes matricu, ir svarīgi zināt arī fokusa plaknes attālumu, jo no tās ir tieši atkarīga arī nobīde starp acīm. [9]

1.2. Perspektīva un projekcija

Vēl viens fundamentāls telpas uztveres pamatprincips ir perspektīva. Vēsturiski tās definīcija arī bija grafisks attēlošanas veids, kas pēc iespējas līdzinās cilvēka redzētajam. Divas galvenās perspektīvas iezīmes ir tas, ka tālāki objekti izskatās mazāki, un ka objekti, kas pagriezti perpendikulāri pret skata virzienu izskatīsies garāki, nekā tie, kas pagriezti tuvāk paralēlei. Turpretī, no matemātikas viedokļa, perspektīvas īpašības seko no tā, ka dabā paralēlas taisnes perspektīvā konverģē uz vienu punktu, nekad nekrustojoties.

Tādējādi, ņemot vērā perspektīvas īpašības, ja ir iespēja attēlā nošķirt kādus objektus vai tā atsevišķas daļas, tad par tā novietojumu telpā var spriest pēc to pozīcijas attēlā, šķietamā formas izkropļojuma un izmēru maiņas dažādās objekta daļās. Piemēram, starp objekta tuvumu attēla centram un tālumu no skatītāja ir sastopama stingra korelācija. Bet, ja divi objekti novietoti uz līdzena seguma, tad tas, kura apakša atrodas augstāk, visdrīzāk ir tālāk.

1.3. Papildinformācija, pieredze

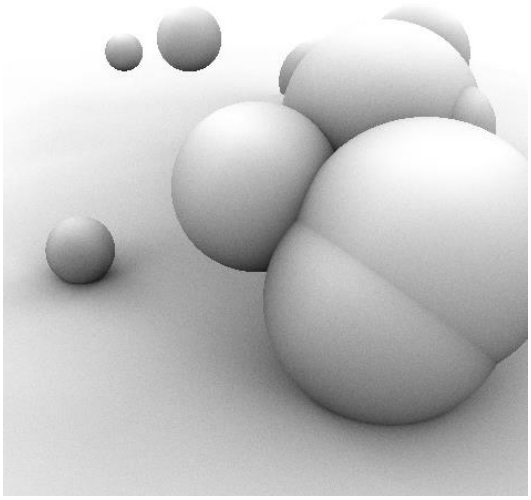
Reti gadās, ka attēlā redzamā informācija dod pilnīgu priekštatu par to, kas tieši tur attēlots. Papildus pašam attēlam ir jābūt arī empīriskām zināšanām par dažādu objektu un vides uzbūvi un izskatu. Ja tādas ir, var veikt pilnvērtīgāku attēla interpretāciju, un aizpildīt trūkstošo informāciju ar minējumu no pieredzes.

Lai spētu šādi interpretēt attēlu, sākumā ir jāspēj identificēt, kas un kur vispār attēlā ir redzams, jeb jāveic objektu atpazīšana un segmentācija. Kad ir saprasts, kas tas ir par objektu, var atsaukties uz iepriekšējo pieredzi un, piemēram, secināt, cik tālu ir objekts, jo ir zināms, cik liels tas ir.

1.4. Kompozīcija

Derīgu informāciju par objektu dziļumu var sniegt arī tas, kurš objekts kuram or priekšā, jeb aizklāj to. Lai gan aizklāšana pati par sevi nedod precīzu dziļuma informāciju, tā var palīdzēt interpretēt attēlu pareizā veidā un izveidot pareizu kompozīciju.

Turklāt, pat ja pats attēlotais objekts ir redzams ar šķietami vienkrāsainu un vienmērīgu virsmu, apkārtējā vidē joprojām ir informācija par precīzāku šī objekta formu un izvietojumu. Atstarotās gaismas daudzums ir atkarīgs no leņķa, kurā tā apstaro objektu [11], vispārējais gaišums konkrētā vietā ir atkarīgs no tā, cik daudz citu objektu ir tam apkārt, bet ēnu forma un novietojums ir tieši atkarīga no objektu savstarpējā telpiskā novietojuma. Piemēram, skat. *1.2. att.*.



1.2. att. *Pēc apgaismojuma intensitātes ir nosakāma kreisās mazās lodes aptuvenā pozīcija*

1.5. Kustība

Kustība kā attēla izmaiņas laikā dziļuma informācijas kontekstā ir līdzīga stereoskopijai, jo tā arī ļauj skatīties no dažādiem skatu punktiem, un ar triangulācijas palīdzību noteikt telpisku informāciju. Bet, atšķirība no stereoskopijas, šajā gadījumā analīzei var izmantot vairāk, nekā divus skatu punktus, un atšķirība starp tiem var būt krietni lielāka. Var nodalīt divus pēc būtības atšķirīgus gadījumus - cita objekta kustību vidē un paša novērotāja kustību. Ja kustās atsevišķa daļa no attēla, tad tas var ļoti uzlabot tikai konkrētās daļas dziļuma informāciju, bet, ja kustas pats novērotājs - visu attēlu.

Kā lielākos trūkumus šai metodei var minēt to, ka tā nebūs pielietojama brīžos, ka kustība nenotiek, un ka vidē var būt vairākas savstarpēji neatkarīgas kustības, kas sarežģī pareizu vides interpretāciju.

1.6. Virtuālās realitātes vides attēlošanas modelis

Virtuālās realitātes ierīces reālistiskas vides ilūziju galvenokārt veido ar četriem pamatprincipiem. Tie ir izolācija no apkārtējās vides, plašs redzes lauks, stereoskopija un vides nekustība attiecībā pret novērotāju. [12] Lai tos nodrošinātu, tiek lietotas speciāli šim uzdevumam izveidotas virtuālās realitātes brilles.

Brilles sastāv no viena vai diviem displejiem, divām lēcām, sensoriem, kas mēra kustību un korpusa. Noslēgts korpuss pilnībā pārklāj abas acis un neļauj tajās nokļūt gaismai no apkārtējās vides vai otras acs, tādējādi nodrošinot izolāciju starp acīm un no apkārtējās vides. Katra acs redz savu displeju, kurš attēlo attiecīgās acs attēlu, kopā izveidojot stereoskopisku attēlu. Lai cilvēka acs spētu nofokusēties uz displeju, kurš atrodas pašās brillēs tikai pāris centimetru attālumā no acīm, un attēls šķistu pietiekami liels, tiek lietotas lēcas ar diezgan lielu kropļojuma koeficientu. [13]

Parasti šādas brilles pašas neveic neatkarīgu darbību, bet ir perifēras ierīces, kas tiek pieslēgtas pie datora, viedtālruņa vai spēļu konsoles. Viedtālruņa gadījumā pašas brilles var arī darboties kā fizisks ietvars bez elektronikas, izmantojot viedtālruņa displeju un sensorus.

Tā kā brilles tiek piestiprinātas pie cilvēka galvas, tad, kustinot galvu, attiecībā pret lietotāja skatu punktu, briļļu rādītais attēls praktiski nemainās, un lietotājam rodas iespaids, ka virtuālā vide griežas un kustas līdzi tam pašam. Tādēļ, lai nodrošinātu virtuālās vides nemainību arī veicot kustības, brillēm ar sensoriem šīs kustības ir jāreģistrē, un ar ļoti mazu aizturi (vēlams zem 20 ms [14]) jāveic attiecīga korekcija redzamajā attēlā. Ja tas netiks darīts, vai arī būs novērojama pārāk liela aizture, tad zudīs reālisma efekts, un bieži vien cilvēkiem īslaicīgi var palikt slikti, parādoties simptomiem, kas līdzīgi šūpes kaitei [15]. Lai sniegtu vēl lielāku reālisma efektu un samazinātu iespējamos negatīvos simptomus, sarežģītākas VR ierīces ar sensoru palīdzību nosaka ne tikai rotāciju, bet arī pozicionālo informāciju. [16]

1.7. Virtuālai realitātei paredzētā satura īpašības

Veidojot saturu, ko var rādīt VR ierīcēs, ir jāņem vērā vairāki papildnosacījumi, salīdzinot ar standarta vizuālajiem medijiem.

Pirmkārt, saturs ir jānodrošina ar pietiekami lielu skata leņķi, lai skatītājs varētu redzēt saturu ne tikai tieši sev priekšā, bet arī perifērajā redzē, pat, ja galva ir pagriezta prom no attēla centra. Praktiski tas nozīmē, ka minimālais skatu leņķis, kas nodrošina apmierinošu satura kvalitāti, ir apmēram 150 grādi. Salīdzinājumam, parastas filmas parasti tiek filmētas ar skatu

leņķi starp 40 un 62 grādiem. [17, 18] Līdz ar to, lai noklātu visu plašo skata laukumu, šādam saturam ir jābūt arī ar lielu izšķirtspēju.

Otrkārt, līdzīgi iepriekšminētajai lietotāja jūtībai uz kustības trūkumu virtuālajā vidē pie attiecīgas kustības realitātē, liekas kustības attēlotajā saturā var lietotājam radīt līdzīgas problēmas. Tādēļ, veidotais saturs ir pēc iespējas jāstabilizē.

2. ATTĒLU APSTRĀDES NEIRONU TĪKLU PAMATPRINCIPI

Šajā nodaļā aplūkoti mākslīgo neironu tīklu pamatprincipi un metodes, kā tos pielietot attēlu apstrādē.

2.1. Mākslīgais neironu tīkls

Mākslīgais neironu tīkls (turpmāk - neironu tīkls) ir viens no galvenajiem rīkiem, ko lieto mašīnmācīšanās nozarē. Pēc uzbūves tas attālināti līdzinās bioloģiskiem neironu tīkliem, kuri darbina dzīvnieku smadzenes, jo abi sastāv no daudziem „neironiem", kas savā starpā dod signālus atkarībā no tiem padotā signāla. Savienojot tīklā pietiekami daudz neironu, ar šādu tīklu ir iespējams atveidot jebkuru patvaļīgu funkciju patvaļīgā definīcijas apgabalā ar patvaļīgu precizitāti [19]. Līdz ar to, pieņemot, ka datoram pietiks resursu šāda tīkla izveidošanai un darbināšanai, neironu tīklu var izmantot jebkuram pielietojumam.

Tipisks mākslīgais neironu tīkls sastāv no vairākiem slāņiem, kur katrs slānis satur neironu kopu 1-3 dimensijās, veic salīdzinoši vienkāršu apstrādi, un tad noteiktā veidā padod savu signālu neironiem nākošajā slānī. Tādā veidā, pirmajā slānī ievietojot ievaddatus, pēdējā slāni tiks izdots vēlamais rezultāts.

Šādi tīkli tiek konstruēti, sākumā izveidojot pašu tīkla struktūru, un tad šo struktūru trenējot ar empīriskiem datiem, kas sniedz informāciju par problēmas risinājumu. Sākotnēji tīkls nedos vēlamo rezultātu, bet, aprēķinot kļūdas lielumu pareizi veidotu tīklu var attiecīgi koriģēt, atvasinot kļūdu uz tīkla parametriem un attiecīgi tos izmainot. Pēc pietiekami daudz korekcijām ar datiem, kas sniedz pietiekami daudz informācijas par problēmas risinājumu, tīkls sāktu dot apmierinošus rezultātus.

2.2. Konvolūciju tīkli

Viens no mākslīgo neironu tīklu paveidiem ir konvolūciju tīkli (CNN). [20] Atšķirībā no tradicionālā jeb pilnsaistes tīkla, kur katrs neirons ir savienots ar visiem neironiem iepriekšējā un nākošajā slānī, konvolūciju tīkli ir savienoti tikai ar tuvākajiem „kaimiņiem" citos slāņos. Šāda struktūra nodrošina vairākas priekšrocības.

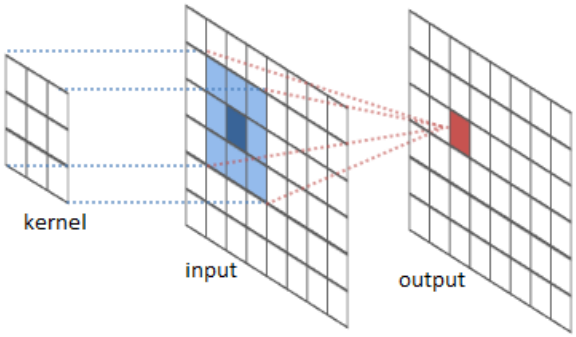
Pirmkārt, atšķirībā no pilnsaistes tīkla, CNN tīklam ar pietiekami daudz slāņiem nobīdes ievaddatu struktūrās nav svarīgas, jo dziļākos slāņos tās pašās izlīdzinās.

Otrkārt, ar konvolūciju palīdzību var panākt, ka, aprēķinot kāda neirona vērtību, tīkls ņems vērā tikai lokālus datus, nevis nesaistītu informāciju otrā ievaddatu kopas galā, līdz ar to būtiski samazinot iespēju, ka tīkls trenēsies no trokšņiem un nejaušām sakritībām.

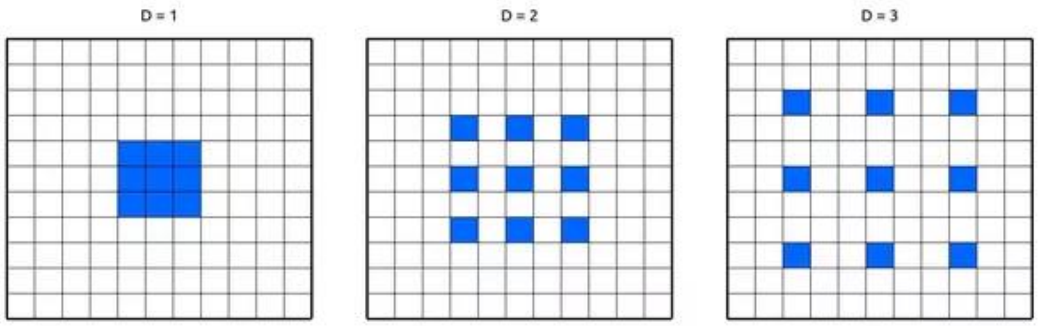
Treškārt, no skaitļošanas viedokļa konvolūciju tīkls tērē daudzkārtīgi mazāk atmiņas un laika, līdz ar to nodrošinot spēju apstrādāt lielizmēra ievaddatus arī ar mūsdienu ierobežotajiem resursiem.

Šo iemeslu dēļ konvolūciju tīkli ir *de facto* risinājums attēlu, video, skaņas un valodas apstrādei un ieteikumu sistēmām. [21]

Konvolūciju tīklam ir 3 galvenie konfigurācijas parametri - kodols, solis un izkliede. Kodols nosaka konvolūcijas izmēru, jeb to, cik daudz attiecīgo neironu no iepriekšējā slāņa tiks apskatīti (Skat. *2.1.att.*), solis nosaka, cik tālu konvolūcijas kodols tiek pārbīdīts katrā solī, un izkliede nosaka atstarpi starp kodola skartajiem elementiem (Skat. *2.2.att.*).



2.2.att. **Konvolūcijas kodols (kernel) nosaka izmantoto neironu skaitu un formu**



2.2. att. **izkliede nosaka atstarpes starp kodola elementiem**

2.3. Attēla ievaddatu interpretācija neironu tīklā

Attēlu apstrādes gadījumā neironu tīkli nodrošina diezgan tiešu sakarību starp rastra attēlu un neironu tīkla ievaddatiem. Parasti, viens pikselis ir viens ievades, bet attēlu ģenerācijas vai transformācijas gadījumā arī izvades neirons. Pirms attēla pikselis tiek ievadīts tīklā, tā vērtības tiek normalizētas uz peldošā punkta skaitli diapazonā 0 līdz 1 vai arī -1 līdz 1. Atkarībā no konkrētās problēmas specifikas, attēls var tikt ievadīts gan krāsainā, gan melnbaltā formātā.

Tomēr, mūsdienu datoru veiktspējas un atmiņas ierobežojumi parasti praktiski ļauj apstrādāt tikai attēlus ar salīdzinoši zemu izšķirtspēju, un tipiska tīkla arhitektūra liek attēla izšķirtspējai būt dalāmai ar 2 kādā pakāpē vai pat konstantai, tādēļ ierastā prakse ir pirms attēla ievades to samazināt līdz kādai noteiktai izšķirtspējai, vai arī apgriezt, ja ir attēla daļas, kuras apstrādāt nav nepieciešams. [22]

2.4. Tīkla dziļums, iezīmju atpazīšana, apkopojošie slāņi

Neirona tīkla vēlamais dziļums, jeb slāņu daudzums, ir problēma, kurai nav vienozīmīgas atbildes. Vēlamais dziļums ir atkarīgs no tā, kāda informācija par attēlu ir nepieciešama, lai varētu veiksmīgi izpildīt prasīto apstrādi. Var gadīties, ka ir nepieciešami elementi gan no seklas, gan no dziļas tīklu struktūras.

Dziļāki tīkli var veiksmīgāk iegūt informāciju no attēla un pārvērst to struktūrā, kas vairs nav dimensionāli atkarīga no paša attēla. un ar šo informāciju veikt sarežģītu attēlu analīzi. Tajos ar konvolūciju palīdzību bildes dimensionālā informācija tiek pārvērsta par iezīmju, struktūru un raksturu kopu, kuru pēc tam var analizēt un izdarīt abstraktus secinājumus par doto attēlu. Mašīnmācīšanās nozari, kas pēta dziļos neironu tīklus, sauc par dziļo mašīnmācīšanos. [23]

Telpiskas informācijas iegūšanas gadījumā vislabāk ir piemēroti tieši dziļi tīkli, jo, kā minēts nodaļā **1.1.**, ir ļoti daudzi veidi, kā šo informāciju var noteikt, no kuriem lielākā daļa ir samērā algoritmiski sarežģīta. Tāpat labāku efektivitāti dziļi tīkli panāk arī iztrūkstošu detaļu aizzīmēšanā, jo tur vajag atpazīt sarežģītas iezīmes, un tās aizpildīt pēc pieredzes. [24] Tas noder arī šī darba kontekstā, jo neliela daļa no ģenerētā kadra nebūs redzama no ievadītā divdimensiju attēla skatupunkta.

Tomēr, lai būtu iespējams veidot dziļus neironu tīklus arī ar apjomīgiem daudzdimensiju ievaddatiem, ņemot vērā, ka dziļiem slāņiem ir nepieciešami daudzi kanāli, dziļākos slāņus ir nepieciešams samazināt izšķirtspējā. To var izdarīt, konvolūciju slānim definējot soli, kas lielāks par 1, vai arī ar apkopojošo slāni (*pooling layer*). Apkopojošais slānis ir slānis, kas paņem n, $n \cdot m$, $n \cdot m \cdot k$ (atkarībā no dimensiju skaita) neironu fragmentus, un agregē tiem kopējo vērtību nākošajā slānī. Vispopulārākais ierobežojošais slānis attēlu apstrādē agregē maksimālo vērtību (*maximum pooling*), jo tādā veidā tiek zaudēts mazāks svarīgas informācijas apjoms.

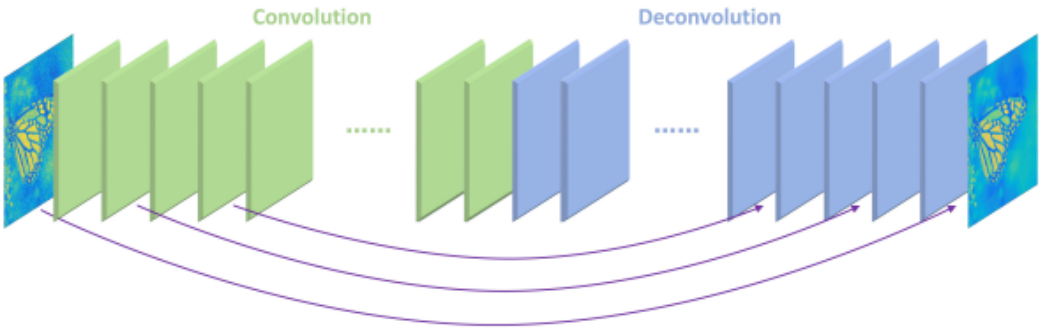
Turpretī seklāki tīkli var veikt tikai virspusēju attēla analīzi, bet satur vairāk precīzas pozicionālas un kompozicionālas informācijas, ir ātrāk un veiksmīgāk trenējami ar mazāku

treniņdatu kopu, kā arī parasti veiktspējas ziņā pārāki par dziļākiem tīkliem. Ja pētāmai problēmai ir nepieciešama liela izšķirtspēja, tad liels tīkla dziļums praktiski var būt neiespējams tehnisku iemeslu dēļ.

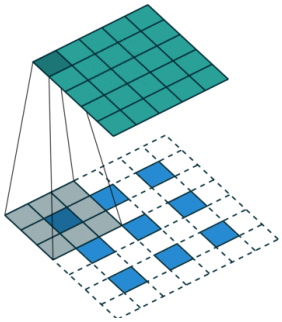
Seklākus tīklus šī darba problēmas risināšanā labāk ir izmantot paša attēla transformācijā, jo tad vēl ir pieejams lielāks precīzas informācijas daudzums par attēla kompozīciju un kopainu. Papildus tam, seklāks tīkls var veikt sekundāru, lokalizētu dziļuma analīzi ar zemāku noteiktību, bet lielāku pozicionālo precizitāti, tādā veidā sniedzot uzlabojumu rezultējošo dziļuma datu precizitātei.

2.5. Izšķirtspējas saglabāšana, pārleciena slāņi, konkatenācija, transponētās un dziļumā atdalāmās konvolūcijas

Lai apvienotu dziļu un seklu tīklu priekšrocības, un dziļa tīkla galā joprojām varētu uzģenerēt attēlu, kurš satur arī sīkas detaļas no orģināla, var veidot tīklu ar *pārleciena* slāņiem, jeb slāņiem, kas ir savienoti nevis ar nākošo, bet gan kādu tīklā tālāk sastopamu slāni (Skat. *2.3. att*). [25] Tas nodrošina iespēju dažādiem datiem plūst cauri tīklam ar dažādu optimālo "apmeklēto" slāņu skaitu. Vienkārši, bet pārāk apjomīgi dati tiek apstrādāti tikai caur pāris slāņiem, bet sarežģītie caur visiem.



2.3. att. **Tipisks neironu tīkls ar pārleciena slāņiem (violetās bultas)**



2.4. att. **transponētā konvolūcija**

Lai tīkla beigu galā būtu iespējams apvienot dziļo slāņu zemas frekvences datus ar augstas frekvences sākuma datiem, sākumā ir nepieciešams palielināt dziļo slāņu izšķirtspēju. To var darīt gan vienkārši atkārtojot datus vairākas reizes katrā izšķirtspējas dimensijā (*upsampling layer*), vai, ja resursi to atļauj, var izmantot transponēto jeb daļas konvolūciju, kas iemācās optimālāko izšķirtspējas palielināšanas metodi trenēšanas gaitā. Tā darbojas pēc principa, ka

zemākās izšķirtspējas ievaddati tiek izretināti ar nullēm starp tiem, un tad augstākas izšķirtspējas slānis tos var aprēķināt ar parastas konvolūcijas metodēm (Skat. *2.3. att*).

Kad abi slāņi ir vienādā izšķirtspējā, tos var apvienot. Ir daudzi veidi, kā to var izdarīt, bet populārākais ir konkatenācija kanālu virzienā. Tad nākošajā slānī var veikt konvolūciju, kas iepriekšējo slāņu datus apvienos pēc trenētā tīkla ieskatiem.

Vēl viens noderīgs konvolūciju veids ir dziļumā atdalāmās konvolūcijas. Tajās, tā vietā, lai katram neironam no konvolūcijas ievada slānī ar katru neironu izvada slānī būtu savs svars, konvolūcija tiek sadalīta divās daļās. Sākumā tiek veikta konvolūcija starp ievada un pagaidu slāni, nemainot dziļuma datus, pēc tam 1x1 konvolūcija starp pagaidu un izejas slāni. Šādā veidā bieži var samazināt trenējamo parametru daudzumu, būtiski nesamazinot efektivitāti. [26]

2.6. Disparitātes, gradienta un dziļuma slāņi

Lai neironu tīklā no disparitātes matricas iegūtu attēlu, ir jāveic otra attēla transformācija ar tā disparitātes matricu, aprēķinot katra pikseļa nobīdi. To var izdarīt, izveidojot jaunu tīkla slāni ar paša definētu funkciju. Tā kā mainīga nobīde nav diferencējama neironu tīkla trenēšanas laikā, tā vietā disparitātes matrica *D* neironu tīklā ir jāveido kā *n* dažādu nobīžu varbūtības, ar varbūtību kopsummu $\sum_n D_{i,j}^n = 1$, kur *i, j* ir pikseļa pozīcija attēlā. Šis varbūtības aprēķinās slānis ar dziļumu *n* un *softmax* aktivācijas funkciju. Tad katras nobīdes varbūtību katram pikselim var reizināt ar sākotnējā attēla *I* attiecīgo pikseli *i,j*, kas nobīdīts par attiecīgās varbūtības piešķirto nobīdi *d:* $I_{i,j}^d = I_{i,j-d}$, un kopā sasummējot visus nobīžu varbūtības reizinājumus $O_{i,j} = \sum_d I_{i,j}^d D_{i,j}^d$, sanāks viens vesels pikselis ar aprēķināto nobīdi katram sākotnējās bildes pikselim. [27]

Viens no būtiskiem attēla kvalitātes rādītājiem ir trokšņu neesamība un vienmērība. To var noteikt, attēlam izmērot gradientu, jeb starpību starp pikseli un tā kaimiņiem. $G_{i,j} = \frac{(I_{i+1,j}-I_{i,j})+(I_{i,j}-I_{i-1,j})+(I_{i,j+1}-I_{i,j})+(I_{i,j}-I_{i,j-1})}{4} = \frac{I_{i+1,j}-I_{i-1,j}+I_{i,j+1}-I_{i,j-1}}{4}$ Tam arī ir var izveidot savu slāni, un tad to var izmantot attēla metriku un zuduma funkcijas aprēķinos. Šajā gadījumā visparastākais aprēķins ir atvasināms, tādēļ nav nepieciešamības meklēt sarežģītāku modeli. [28]

Lai viegli varētu aprēķināt metrikas ar disparitātes datiem, tās ir jāpārvērš no nobīžu varbūtību saraksta par matricu ar pašām nobīdēm, nepārvēršot tās bildē. Katram pikselim to var

izdarīt, sareizinot katru varbūtību ar attiecīgo nobīdi, un reizinājumus saskaitot kopā. Rezultātā iegūst parastu disparitātes matricu $D'_{i,j} = \sum_d (d \cdot D^d_{i,j})$.

2.7. Vērtības zuduma funkcija

Lai neironu tīklu varētu veiksmīgi trenēt, ir jāspēj noteikt, cik tālu no vēlamā rezultāta tas pašreiz atrodas, un atvasināt šo kļūdu, lai noskaidrotu, cik tālu kurā virzienā pielabot parametrus. Tam ir paredzēta vērtības zuduma funkcija. Ir izpētīts, ka darbam ar attēliem, kur ir svarīgi atsevišķi pikseļi, labākā kļūdas aprēķina funkcija ir vidējā absolūtā kļūda (L1). [29]

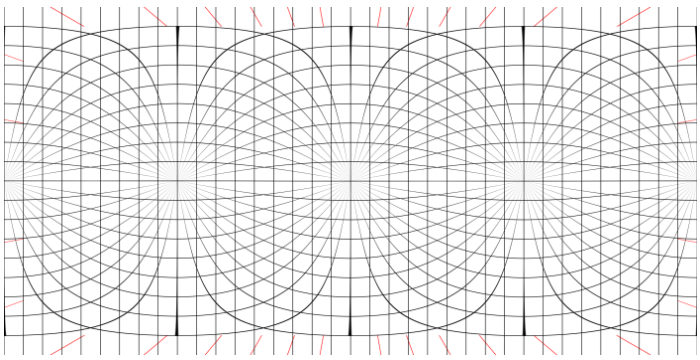
Galvenokārt šādu kļūdu aprēķina attiecībā pret tiešo gaidāmo rezultātu, bet ir arī specifiskāki gadījumi.

Izmantojot iepriekš definētos gradienta un dziļuma slāņus, var izveidot vērtības zuduma aprēķinu, kas no disparitātes matricas aprēķina pašu disparitāti, tad tai aprēķina gradientu, un reizina to ar izveidotā attēla apgrieztu melnbaltu gradientu. Šis rezultāts tad tiek pielīdzināts nullēm. $Kļūda = (1 - G_{melnbalts\ attēls}) \cdot G_D$, Pēc būtības, šis kļūdas aprēķins nosaka, ka vietās, kur attēls ir vienmērīgs, vienmērīgam ir jābūt arī dziļumam. Līdz ar to tiek panākta lieku trokšņu noņemšana. [28]

2.8. Virtuālās realitātes satura attēlojums neironu tīklā

Lielākā daļa virtuālās realitātes video materiālu ir saglabāta ģeogrāfiskajā projekcijā - attēla x un y koordināta ir tieši proporcionāla horizontālajiem un vertikālajiem leņķim λ un ϕ. Tas ir matemātiski vienkāršs modelis, ar kuru var attēlot arī visu 360 grādu sfēru un ļaut atrast projekciju noteiktā virzienā, vienkārši translējot daļu attēla, bet tā projekcija uz plaknes tomēr ir ar attēla kropļojumiem. [30]

Konvolūciju tīkli labi strādā ar lineārām transformācijām un tās kompensē, bet atpazīt nelineāru transformāciju bez lokāliem neironu tīklu svariem prasa daudz sarežģītāku trenēšanu. Turklāt, strādājot ar disparitātes matricām, jebkura nobīde vertikālā virzienā var pilnīgi izbojāt iespēju atrast atbilstošo pikseli blakus



2.5. att. **Horizontālās līnijas ģeogrāfiskajā projekcijā gandrīz vienmēr nav horizontālas**

attēlā, jo tas ir nobīdīts arī vertikāli (Skat. *2.5. att.*). Tādēļ, pirms apstrādāt VR video materiālu, to ir labāk pārvērst rektilineārā projekcijā, kur jebkura taisne arī izskatās kā taisne (Skat. *2.6. att.*) [31].

$$x = \frac{cos\phi\, sin(\lambda - \lambda_0)}{cos(c)}$$

$$y = \frac{cos\phi_1 sin\phi - sin\phi_1 cos\phi cos(\lambda - \lambda_0)}{cos(c)} \qquad cos(c) = sin\phi_1 sin\phi + cos\phi_1 cos\phi cos(\lambda - \lambda_0)$$

2.6. att. **Rektilineārās projekcijas aprēķins no ģeogrāfiskās projekcijas, λ_0 un ϕ_1 ir projekcijas rotācijas centrs**

Kad attēls ir uzģenerēts, tā lietošanai VR ierīcēs to vajag deformēt atpakaļ uz ģeogrāfisko projekciju (Skat. *2.7. att.*) [31].

$$\phi = sin^{-1}\left(cos(c)sin\phi_1 + \frac{ysinccos\phi_1}{\rho}\right)$$

$$\lambda = \lambda_0 + tan^{-1}\left(\frac{xsin(c)}{\rho cos\phi_1 cos(c) - ysin\phi_1 sin(c)}\right)$$

where, $\rho = \sqrt{x^2 + y^2}$ and $c = tan^{-1}\rho$

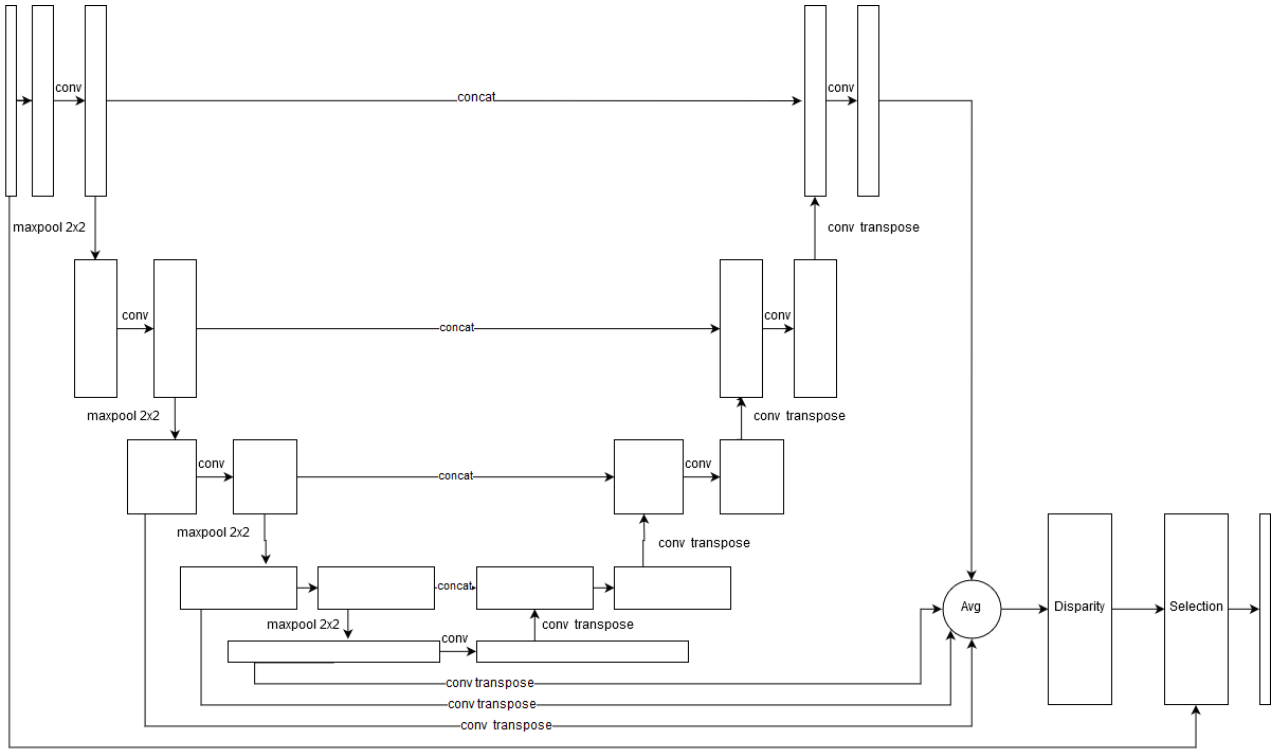
2.7. att. **Ģeogrāfiskās projekcijas aprēķins no rektilineārās projekcijas, λ_0 un ϕ_1 ir projekcijas rotācijas centrs**

3. RISINĀJUMA MODELIS

Risinājuma modelis ir balstīts uz iepriekšējā pētījumā izveidotu neironu tīklu *wnet* [28], kurš ģenerē dziļuma datus, kā treniņdatus izmantojot stereoskopiskus video materiālus. Iepriekšminētais pētījums savukārt ir balstīts uz vecāku pētījumu par stereoskopisku video ģenerāciju no parastiem video materiāliem *deep3d*. [27]

Darba autors izpētīja šos modeļus, un tad mēģināja tos pilnveidot un pielāgot gan vispārīgi, gan ņemot vērā VR video mediju ģenerācijas specifiku.

3.1. Modeļa sākotnējā struktūra



3.1. att. **Risinājuma sākotnējā struktūra**

Darba autors kā sākotnējo modeļa struktūru izvēlējās *wnet* modeli, samazinot to uz pusi iezīmju dimensijā, jo šī darba specifika neprasa ģenerēt abus stereoskopiskos kadrus, bet gan tikai vienu (Skat. *3.1. att.*). Ar daudziem paralēliem savienojumiem šis modelis nodrošina to, ka visa informācija, kas atrodas tīklā, arī nonāk līdz beigām. Pirmkārt, Tajā ir pieci dažādi līmeņi, katrs no kuriem ir divas reizes mazāks par iepriekšējo abās datu dimensijās, bet divreiz lielāks iezīmju dimensijā. Šie līmeņi ir savā starpā savienoti ar apkopojošajiem slāņiem, samazinot datu dimensiju, un transponētām konvolūcijām, palielinot datu dimensiju. Starp vienādu līmeņu izmēra slāņiem veidotas parastas 3x3 konvolūcijas, ja slāņi ir secīgi, vai

konkatenācija, ja slāņi ir savienoti ar pārlecienu. Beigās tiek aprēķināts vidējais aritmētiskais no pēdējās konvolūcijas un trīs dziļākajiem apkopojošajiem slāņiem, kuru savukārt izmanto *softamax* slānis, lai aprēķinātu katras nobīdes iespējamību katram pikselim. *Selection* slānis šo nobīdi izpilda uz ieejas datiem, un tad atdod rezultātu.
Tālāk aprakstītas izskatītās un izmēģinātās izmaiņas šajā modelī.

3.2. Ieejas datu pirmsapstrāde

Tā kā VR video materiāliem ir jābūt platleņķa formātā, un platleņķa bildes nevar attēlot plaknē bez optiskiem kropļojumiem (Skat. **2.8.**), tad, lai šos kropļojumus novērstu, un varētu iegūt viendabīgu attēlu, kur taisnas līnijas arī izskatās taisnas, ir jāveic izliekuma korekcija. Lai veiktu šo korekciju, tika izveidota funkcionalitāte, kas aprēķina attēla transformācijas kartes, ar kurām ātri transformēt ģeogrāfiskas projekcijas attēlus rektilineārajā projekcijā un atpakaļ.

3.3. Modeļa struktūras variācijas

Pētot modeļa sākotnējās struktūras aktivācijas uz disparitātes matricu, darba autors konstatēja, ka dziļie slāņi caur pārleciena transponēto konvolūciju uz disparitātes matricu pa taisno izdod diezgan maz aktivāciju. Tādēļ, darba autors izmēģināja divus modeļus, kuros ir izņemti tiešie pārlecieni uz disparitātes aprēķinošo slāni no attiecīgi pirmā visdziļākā un no trim visdziļākajiem slāņiem.

Turpinot pētīt modeļa daļu, kas padod aktivācijas no dziļajiem slāņiem zemā izšķirtspējā atpakaļ uz pilnas izšķirtspējas disparitātes aprēķina slāņiem, darba autors secināja, ka aktivāciju daļas no katras izšķirtspējas slāņiem ir vienādas un fiksētas, jo no tām tiek rēķināts aritmētiskais vidējais. Lai dotu iespēju neironu tīklam pašam izvēlēties slāņu aktivācijas proporciju, tika izveidoti četri trenējami skalāri neironi, kuri regulē dažādo slāņu aktivāciju proporciju.

Kā alternatīva iepriekšējai dažādu līmeņu aktivācijas apvienošanas metodei tika izmēģināta visu slāņu konkatenācija, aiz kuras tika pievienots vēl viens konvolūciju slānis ar 1x1 kodolu, kas tādējādi nevarētu izmainīt izšķirtspējas datus, bet var precīzi apvienot visu slāņu iezīmju dimensijas.

Tika apskatīta arī alternatīva metode, kā veikt slāņa izšķirtspējas palielināšanu. Transponēta konvolūcija tika aizstāta ar 1x1 kodola konvolūciju, kurai seko mērogošanas slānis, kurš katru neironu aktivāciju ieejas slānī pārvērš par NxN identiskām aktivācijām izejā.

Tādā veidā var panākt, ka dziļuma dati ir mērogoti vienmērīgi, nevis ar nejauši uztrenētiem trokšņiem, kas varētu veidoties, ja tiktu izmantota transponēta konvolūcija ar lielu kodola izmēru.

Lai iepriekš izveidotā zemas izšķirtspējas informācija rezultējošā attēla neizraisītu fragmentēšanos un asas robežas, tika izveidots vēl viens konvolūciju slānis ar izkliedētu kodolu, kurš noklātu un interpolētu blakus esošās vērtības no zemās izšķirtspējas slāņa. Tika izmēģināti divi kodolu izkliedētības varianti, 5x5 kodols ar izkliedētību 5 un 9x9 kodols ar izkliedētību 3 izšķirtspējas palielināšanai 16 reizes, 5x5 kodols ar izkliedētību 3 un 7x7 kodols ar izkliedētību 2 palielināšanai 8 reizes, un 3x3 vai 5x5 kodols bez izkliedētības palielināšanai 4 reizes.

Ņemot vērā neapmierinošo augstas frekvences detaļu dziļuma noteikšanu, modeļa slāņiem, kuri bija pilnā izšķirtspējā, tika divas reizes palielināta iezīmju dimensija. Tajos pašos nolūkos pirmajā slānī kā ieejas dati arī tika pievienots arī melnbaltas ieejas bildes gradients. Darba autors hipotizēja, ka tas neironu tīklam varētu ļaut vieglāk attēlā atrast pozīcijas, kuras potenciāli trīs dimensiju telpā var būt šķautnes.

Kā vēl viens optimizējams hiperparametrs tika pētīts slāņu iezīmju skaita reizinātājs. Tas ir skaitlis, ar kuru dalās visu slāņu iezīmju dimensijas garumi. Līdz ar to, mainot šo skaitli, proporcionāli mainās visa neirona tīkla izmērs. Tika pārbaudīti četri lielumi - 8, 16 un 32.

3.4. Modeļa pētītā papildfunkcionalitāte

Tika pētīts, vai modelis spēs darboties, ja tam pēc pieprasījuma ir ne tikai jāģenerē labās acs attēls no kreisās acs attēla, bet arī kreisās acs attēls no labās acs. Par šādu ģenerācijas variantu neironu tīklam tiek paziņots ar skalāru parametru, kurš var būt 1 (ģenerēt labo aci) vai -1 (ģenerēt kreiso aci). Tas tika savienots ar tīkla pirmo slāni, vai arī ar pēdējo slāni pirms disparitāšu varbūtības aprēķināšanas.

Vēl viens piedāvāts papildu parametrs, kurš tiek ievadīts kopā ar ieejas attēlu ir skatu platuma vērtība. Sākotnēji tā tika ieviesta, lai kompensētu datu augmentācijas radītās horizontālās rotācijas sekas (Skat. **4.2.**), tomēr tā potenciālais pielietojums ir krietni plašāks - gluži kā parastās brilles, izmantojot VR ierīces, tās ir jānoregulē cilvēkam atbilstošā acu platumā. [32] Ar šādu parametru varētu regulēt arī ģenerācijas radīto atstarpi starp skatu punktiem, lai tā tieši atbilstu cilvēka acu platumam.

3.5. Tehniskie ierobežojumi

Datoram, uz kura tika veikta tīklu trenēšana un novērtēšana ir 16 GB operatīvās atmiņas, 6 kodolu *Intel Core i7-8700K* procesors un *Nvidia Geforce GTX 980 Ti 6 GB* videokarte. Viens no lielākajiem ierobežojumiem ir video kartes atmiņas daudzums, jo šis modelis jau sākotnēji aizņem aptuveni 5 GB video atmiņas.

3.6. Implementācija

Modelis ir veidots augsta līmeņa mašīnmācīšnās ietvarā *Keras*[1], kurš konfigurēts, lai izmantotu zema līmeņa ietvaru *Tensorflow*[2]. Izmantotais dators strādā ar Windows 10 operētājsistēmu.

Kods[3] ir rakstīts *Python* 3 programmēšanas valodā, un palaists uz *Anaconda*[4] datu zinātnes platformas. Daļa tīkla veidošanas, trenēšanas un diagnostikas koda ir aizgūta no *wnet* projekta, bet darba autors tajā ir veicis apjomīgas izmaiņas (Skat. **3.2.-3.4, 4.2., 4.3.**). Papildus kodam, kas darbina pašu neironu tīklu, darba autors uzrakstīja arī funkcionalitāti, kas veic video materiālu pirmsapstrādi un diagnostikas rīkus neironu tīklu un attēlu projekciju pētīšanai.

Lai apstrādātu video failus un bildes tika izmantota *OpenCV*[5] datorredzes bibliotēka, bet lai sekotu līdzi tīklu trenēšanai un pētītu tās rezultātus, tika lietoti *Jupyter Notebook*[6] un *TensorBoard* rīki.

3.7. Modeļa pielietošana

Pēc trenēšanas pabeigšanas modeļa uzdevums ir ģenerēt stereoskopiskus video materiālus un attēlus no monoskopiska video materiāla vai attēla. Ja datu pirmsapstrādē tiks veikta optisko kropļojumu korekcija, bet pēcapstrāde šo kropļojumu atkārtota pielietošana, modeli varēs lietot arī VR materiālu ģenerācijai.

[1] https://keras.io/
[2] https://www.tensorflow.org/
[3] https://bitbucket.org/karlis383/vrnet/
[4] https://anaconda.org/
[5] https://opencv.org/
[6] http://jupyter.org/

Tā kā modelis ir pilnībā veidots no konvolūciju slāņiem, tam nav stingri izšķirtspējas ierobežojumi pēc modeļa uztrenēšanas, izņemot praktiskus atmiņas apjoma ierobežojumus, un to, ka izšķirtspējai abos virzienos jādalās ar 16. Var gan izrādīties, ka lielākā izšķirtspējā ar pašreizējām disparitātes vērtībām nepietiek, bet šīs vērtības var arī mērogot, zaudējot procentuāli nelielu precizitātes daudzumu.

4. TĪKLU TRENĒŠANA

4.1. Izmantotie treniņdati

Kā treniņdati ir izmantoti stereoskopiski VR video ģeogrāfiskajā projekcijā un formātā, kur kreisās acs attēls atrodas kreisajā pusē, bet labās - labajā. Šis formāts atbilst Google izveidotajam VR180 standartam un ir ieteikts kā viens no veidiem, kā saglabāt VR video. [33]

Treniņdati ir iegūti no dažādām video koplietošanas platformām un ar VR video veidošanu saistītām mājaslapām, kurās bija pieejami paraugi. Treniņdati netika speciāli izlasīti. Tajos tika attēlos dažāds saturs - dabas skati, izbraucieni ar dažādiem transportlīdzekļiem, auto izstādes, koncerti un citi.

Diemžēl, stereoskopiska VR video zemās popularitātes un ierobežoto ielādēšanas iespēju pareizā formātā dēļ, kopā tika atrasti tikai 19 piemēroti video materiāli ar kopējo garumu 51 minūte. Lai no tik maza treniņdatu apjoma varētu iegūt apmierinošus rezultātus, darba autors pievērsa īpašu uzmanību šo datu augmentācijai un pirmsapstrādei, kā arī izmēģināja sākumā trenēt tīklu ar 11 parastām 3d filmām, un tikai beigās pielabot tīklu ar VR video paraugiem, kuros ir kvalitatīvāka stereoskopijas informācija un lielāks skata leņķis.

4.2. Datu augmentācija

Tā kā telpas uztvere ir lielā mērā simetriska pret vertikālo asi, atsevišķi video kadri tika apgriezti spoguļattēlā, kreisās acs attēlu padarot par labās acs attēlu, un otrādi. Darba autors uzskata, ka tas varētu nodrošināt lielāku tīkla vienmērību.

Ģeogrāfiskajā projekcijā, kurā kodēti treniņdati, pikseļa pozīcija x un y attēlā tieši atbilst skata leņķim λ un ϕ no skata centra. Līdz ar to, vienmērīga attēla translācija šajā projekcijā nodrošina lineārās projekcijas rotāciju ap skatu punktu. Tas tika izmantots, veidojot treniņdatus, no katra attēla kadra izveidojot piecus - vienu bez rotācijas, un četrus rotētus 30° uz abām pusēm gan vertikāli, gan horizontāli. Lai gan abu acs projekciju rotācija pa horizontālo aci nozīmētu, ka šķietamais platums starp acīm mainās, modelī tika izveidots papildu ieejas parametrs, kurš kompensē šo šķietamo platumu. Taisna vai vertikāli rotēta attēla gadījumā šī ieejas vērtība būtu 1, bet horizontālas rotācijas gadījumā cos(30°).

4.3. Treniņdatu pirmsapstrāde

Lai pirms katras trenēšanas lieki netērētu laiku, konvertējot video datus, pirmsapstrāde tika izveidota kā neatkarīgs posms, kas strādā tikai tad, ja ir parādījies kāds jauns video materiāls.

Video kadriem tika veikta optiskā kropļojuma korekcija (skat. **3.2.**), tie tika apgriezti un mērogoti zemākā izšķirtspējā, un daļai no tiem tika veiktas iepriekšējā apakšnodaļā minētās datu augmentācijas.

Lai treniņdatos nebūtu lieki kadru dublikāti, katram augmentācijas paveidam tika piešķirts katrs desmitais video kadrs. Tādējādi, katrs ieejas video kadrs tika apstrādāts, bet līdzīgie kadri tika transformēti dažādās formās. Lai izvairītos no kadru dublikātiem arī stacionāros video, kur ilgstoši tiek rādīta viena un tā pati aina, katrs kadrs, kurš piešķirts noteiktai augmentācijai, pēc transformācijas tika salīdzināts ar iepriekšējo šī tipa augmentācijas kadru, izmantojot strukturālās līdzības indeksu [34]. Ja šo kadru līdzība pārsniedza 0.85, kadrs tika izlaists. Šādi tika izlaisti aptuveni 40% no treniņdatu kadriem.

Lai trenējot tīklu tiktu nodrošināta vismaz daļēju nejaušību kadru secībā, trenējot tīklus paralēli tika atvērti 6 video materiāli, un katrā trenēšanas solī nejauši tika izvēlēts un lasīts viens no tiem. Kad kāds video beidzās, tā vietā tika nejauši izvēlēts cits tajā brīdī neatvērts video no datu kopas.

4.4. Trenēšanas parametri

Treniņdatu partijas izmērs (*Batch size*) - palielinot šo parametru, var iegūt ātrāku datu apstrādes ātrumu, kā arī trenēšanas stabilitāti, jo modelī izmantots *batch normalization,* kas normalizē aktivāciju mērogu attiecībā pret visiem paraugiem šajā partijā. Tomēr, lielāka partija arī prasa vairāk atmiņas, tādēļ šī darba gadījumā maksimālais partijas izmērs, kas nekad neradīja atmiņas problēmas, bija 5.

Pēc ievaddatu kropļojumu korekcijas dabā paralēlās līnijas kļūst paralēlas arī attēlā, bet objekti attēla malās joprojām ir izstiepti un neproporcionāli tiem, kas atrodas centrā, savukārt centrālie objekti kļūst ļoti mazi. Turklāt, rektilineārajā projekcijā nemaz nav iespējams attēlot pilnu 180 grādu skatu, jo projekcijai perpendikulārie stari diverģē uz bezgalību. Tādēļ ir nepieciešams šo projekciju "apgriezt", parādot tikai noteiktu centrālo daļu. Bet, ja tomēr ir nepieciešamība pēc lielāka skata laukuma, tad ir jāveic vairākas bilžu ģenerācijas. Tika izmēģināti vairāki apgriešanas varianti - 90°, 105° un 120°.

Datu augmentācijai, kas veic skata rotāciju, var mainīt pagrieziena leņķi. Tika izvēlēts 30° pagrieziena leņķis, jo, izmantojot šādu leņķi kopā ar 90° skata leņķi, tiek noklāts 150° liels laukums, kas atstāj pašas attēla malas, kur papildu ģeogrāfiskās projekcijas kropļojumam ir pamanāms arī optiskais kropļojums, kas radies no nepilnīgām lēcām [35]. Šo kropļojumu ir ļoti grūti novērst, jo tā daudzums un specifika dažādiem video atšķiras.

Atsaucoties uz [29], kā *dropout* tika izmēģinātas vērtības 0.4 un 0.8.

Ievaddatu izšķirtspēja - Modelī ir četri secīgi apkopojošie slāņi, no kuriem katrs izšķirtspēju samazina divas reizes abās dimensijās, un industrijas standarta video materiāla garuma pret platumu attiecība ir 16:9. Tādēļ, izšķirtspējas augstumam ir jādalās ar $9 \cdot 2^4 = 144$, bet puses jeb vienas acs kadra platumam jābūt 8/9 no tā augstuma. No tā seko, ka minimālā izšķirtspēja vienas acs kadram ir 128×144. Tomēr, lai varētu kvalitatīvāk novērtēt uzģenerētos kadrus, pēc sākotnējiem testiem tika lietota tikai 256×288 izšķirtspēja.

4.5. Trenēšanas un validācijas datu sadalījums

Sākotnēji treniņdatu un validācijas kopas bija nodalītas attiecībā 3:1, mēģinot abās iekļaut pēc iespējas dažādus video materiālus. Taču, mēģinot samazināt *overfitting,* tika izveidots krusteniskās validācijas modelis. Tajā treniņdatu un validācijas kopas tika sadalītas attiecībā 9:1, bet abas kopas ir ņemtas no vienas kopīgas video materiālu kopas, un katrā no 10 trenēšanas etapiem validācijas kopa bija atšķirīga no visām iepriekšējām - pēc 10 etapiem visi dati bija validācijas kopā tieši vienreiz.

4.6. Optimizātors, vērtības zuduma funkcija un kļūdu aprēķins

Kā tīkla trenēšanas optimizātors tika izvēlēts *Adam optimizer*, jo tas diezgan labi optimizē tīklu, nepieprasot precīzi pielāgot tā parametrus [35].

Kā vērtības zuduma funkcija izmantota vidējā absolūtā kļūda, jo tīklos, kuros ir svarīga augstas frekvences informācija, tā strādā labāk (Skat **2.7.**).

Kļūdas tika aprēķinātas, salīdzinot izejas attēlu ar treniņdatu referenci, un veidojot dziļuma datu un bildes gradientu salīdzinājumu (Skat **2.7.**), ar svariem 1 pašai bildei un 0.01, 0.001 un 0.0001 gradientu kļūdai.

Tika izmēģināta arī kļūdu funkcija, kas salīdzina rezultējošā attēla gradientu ar treniņdatu references gradientu, bet darba autors nevarēja to izveidot vai pielietot tā, lai trenējot kļūda nediverģētu uz bezgalību.

4.7. Trenēšanas ilgums

Lai testētu kādas izmaiņas potenciālu, tika izmantoti īsi 30-60 minūšu treniņi, un, ja konkrētā izmaiņa neradīja acīmredzamu negatīvu efektu, to trenēja 2-4 stundas.

REZULTĀTI UN DISKUSIJA

Modeļa un trenēšanas mainīgo ietekme uz rezultātu

Dati verificēti ar vidējo absolūto kļūdu pret 500 nesecīgiem kadriem no 3 pēc satura un veida atšķirīgiem video, kurus tīkls iepriekš nav redzējis. Ne-VR 3D filmu gadījumā verificēts pret 2000 nesecīgiem kadriem no divām tīklam neredzētām 3D filmām (bez titriem). Attēli tika pārbaudīti arī ar autostereogrammu, darba autoram mēģinot tajos novērtēt stereoskopijas efekta kvalitāti.

1. tabula

Dažādu modeļu trenēšanas rezultāti

Variācija	**Verifikācijas kļūda**	**Kvalitatīvs autora novērtējums**	**Piezīmes**
Treniņdatu variācijas (ar visveiksmīgāko modeli un trenēšanas parametriem) 90° skata leņķis, ja nav norādīts citādāk			
VR video bez projekcijas korekcijas	0,0434	Ārpus attēla vidus mazas transformācijas	VR ierīces telpiskums būs izkropļots
VR video ar projekcijas korekciju	**0,0637**	Kvalitatīvs telpiskums, maz izplūdusi bilde	
VR video ar projekcijas korekciju 120°	0,0760	Centrālie objekti kļūst pārāk mazi, pazūd detaļas, nedaudz izplūdis.	
VR video ar projekcijas korekciju un datu augmentāciju	0,0690	Kvalitatīvs telpiskums, bet izplūdusi bilde	
VR video ar projekcijas korekciju un datu augmentāciju 105°	0,0752	Skaidri redzami optiskie kropļojumi gar pašu skata malu	
3D filmu treniņdati	**0,0581**	Mazs, bet kvalitatīvs stereoskopijas efekts, ir jāmērogo disparitātes	Vērtēts pret VR video datiem

3D filmu treniņdati	**0,0201**	Atbilst paraugam, bet mazs stereoskopijas efekts	Vērtēts pret 3D filmu datiem
3D filmu treniņdati + VR video ar optiskā kropļojuma korekciju un datu augmentāciju	0,0889	Atšķiras disparitāšu mērogs un fokuss, tāpēc rezultējošā bilde ir ļoti izplūdusi, ir jāmērogo disparitātes	
3D filmu treniņdati + VR video ar optiskā kropļojuma korekciju	0,0639	Atšķiras disparitāšu mērogs un fokuss, tāpēc rezultējošā bilde ir ļoti izplūdusi	
Modeļa variācijas (treniņdati - VR video ar optiskā kropļojuma korekciju un datu augmentāciju, visveiksmīgākie trenēšanas parametri, ja nav norādīts citādāk)			
Sākotnējais modelis	0,0807	Izplūdis attēls ar nodalāmiem zemas frekvences fragmentiem	Skat. 1. pielikumu
Izņemts visdziļākais slānis	0,121	Mazāks, nevienmērīgāks telpiskuma efekts	Izmantota novecojusi datu kopa, bet potenciāls tāpat nav saskatāms
Izņemti 3 visdziļākie slāņi	0,152	Haotisks dziļumu sadalījums	Izmantota novecojusi datu kopa, bet potenciāls tāpat nav saskatāms
4 skalāri neironi, kuri nosaka dažādu slāņu ietekmi uz disparitāti	-	-	Tehnisku iemeslu dēļ šie parametri netrenējās
Alternatīvā izšķirtspējas palielināšana 1x1 conv + upsample	0,0618	Izplūdis, nevienmērīgs attēls	Skat. 1. pielikumu
Papildu zemo izšķirtspēju interpolācija - 5x5 kodoli, 5, 3, 1 izkliedētība	0,0759	Krietni mazāk izplūdis, vienmērīgāks attēls	
Papildu zemo izšķirtspēju interpolācija - 9x9, 7x7 un 5x5 kodoli, 3, 2, 1 izkliedētība	0,0705	Krietni mazāk izplūdis, vienmērīgāks attēls	Skat. 1. pielikumu

Melnbalts ieejas attēla gradients klāt ieejas datiem	**0,0637**	Neliels izplūdums, bet laba kvalitāte	Sevī ietver arī alternatīvo izšķirtspējas palielināšanu un interpolāciju 9x9…5x5
Slāņu reizinātājs 8	0,1215		
Slāņu reizinātājs 32	0,0691	-	Video kartei nepietika atmiņas lai šo modeli atkārtotu ar citiem
Parametrs, ka jāģenerē pretējais skats (pievienots tīkla sākumā)	0,1101	Haotiskāka bilde	Novecojuši dati
Parametrs, ka jāģenerē pretējais skats (pievienots tīkla beigās)	0,0901	Nav saskatāmu izmaiņu	Novecojuši dati
Trenēšanas variācijas (treniņdati - VR video ar optiskā kropļojuma korekciju un datu augmentāciju, visveiksmīgākie modeļa parametri, ja nav norādīts citādāk)			
256x288 izšķirtspēja	**0,0637**		
512x576 izšķirtspēja	0,0638	Mazs stereo efekts	Nav trenēts ar šo izšķirtspēju
512x576 izšķirtspēja, disparitātes mērogs x2	0,0629	Novērojamas nelielas neregularitātēs	Nav trenēts ar šo izšķirtspēju
Dropout 0,8	0,0781		
Dropout 0,4	**0,0637**		Standarta gadījums, pēc tam izmantots visiem
Sekundārās kļūdas koeficients 0,01	0,0797	Ļoti izplūdusi bilde	
Sekundārās kļūdas koeficients 0,001	**0,0637**		Standarta gadījums, pēc tam izmantots visiem
Sekundārās kļūdas koeficients 0,0001	0,0814	Saraustīta dziļuma informācija	
Bez sekundārās kļūdas	0,0766	Saraustīta dziļuma informācija	

Veiktspēja

Tīkla trenēšana līdz labākajiem rezultātiem aizņēma līdz 3 stundām, 10 stundu garu video materiālu pārveidošana trenējamā formā aizņēma aptuveni 2 stundas. 256x288 izšķirtspējā, apstrādāt vienu attēlu prasa vidēji 100-120 milisekundes. Informāciju par datoru skatīt apakšnodaļā **3.5.**.

Salīdzinājums ar citu autoru darbiem

Diemžēl, darba autoram neizdevās veikt tiešu salīdzinājumu ar citiem darbiem, jo *wnet* modelis tika veidots ar mērķi ģenerēt tieši dziļuma datus, un nespēj darboties bez abu stereoskopisko skatu punktu klātbūtnes, *deep3d* modeli darba autoram neizdevās nokompilēt tehnisku iemeslu dēļ, un citi pētījumi netika apskatīti. Tomēr, *deep3d* autori apgalvoja, ka *deep3d* risinājumam vidējā absolūtā kļūda ir 0,0687 [27]. Ir gan jāņem vērā, ka šo rezultātu nevar tieši salīdzināt ar šī darba rezultātiem, jo, iespējams, ka testētās datu kopas atšķiras pēc sarežģītības.

Kvalitatīvs vērtējums

Lielākā problēma risinājuma attīstīšanas gaitā bija izplūdis attēls. Tas nozīmē, ka tīkls *softmax* slānī nemācēja viennozīmīgi klasificēt, cik liela nobīde konkrētajam pikselim ir nepieciešama. Tam par iemeslu var būt gandrīz jebkas, kas var traucēt tīkla darbību, piemēram, datu trūkums, neviennozīmīgi dati, nederīgs modelis. Tomēr, izplūduma samazināšanās bija labs rādītājs tam, ka kāda izmaiņa pozitīvi ietekmē tīkla darbību. Šāds samazinājums bija skaidri novērojams gan pie datu augmentācijas, apstrādes un padeves uzlabojumiem, gan pie modeļa slāņu apvienošanas struktūras izmaiņām, gan arī visskaidrāk, sākot izmantot daudzveidīgākus 3D filmu datus.

Modelim kļūstot efektīvākam, mainījās arī prasības pret datiem - ja sākumā horizontālās rotācijas datu augmentācija palīdzēja sasniegt labāku rezultātu, jo tā trīskāršoja datu kopas apjomu, tad tagad tā ieviesa lieku izplūdumu, un bija izdevīgāk to atslēgt.

Ļoti uzmanīgi tika pētīta arī dažādu izšķirtspēju datu apvienošana atpakaļ vienā disparitātes matricā, jo tajā tika novērotas neregularitātes, kas šķietami tika izraisītas apvienojot zemas frekvences datus bez interpolācijas. Pēc vairākām izmēģinātām alternatīvām tika atrasts

risinājums, kas joprojām izmanto zemas frekvences datus, bet tos pienācīgi interpolē (Skat. 1. pielikumu).

Visos novērojumos un piemēros izmantoti attēli, kuriem līdzīgi nav sastopami treniņdatos.

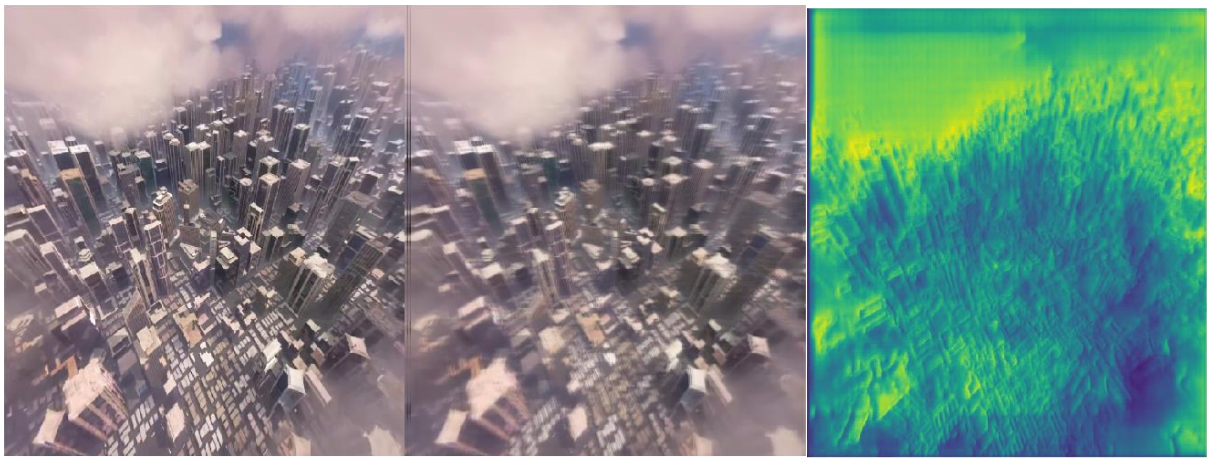
Labākais risinājums

Tīkla pārtrenēšana no 3D filmu datiem uz VR video datiem pagaidām nesniedza vēlamus rezultātus, tādēļ pašlaik labākais risinājums ir 3D filmu datu kopa ar modeli, kuram ir disparitātes slāņu apvienošana ar 1x1 konvolūciju, apkopojošais slānis un izkliedēta 9x9-5x5 konvolūcija interpolācijai, kā arī papildus ieejas vērtība melnbalta gradienta veidā.

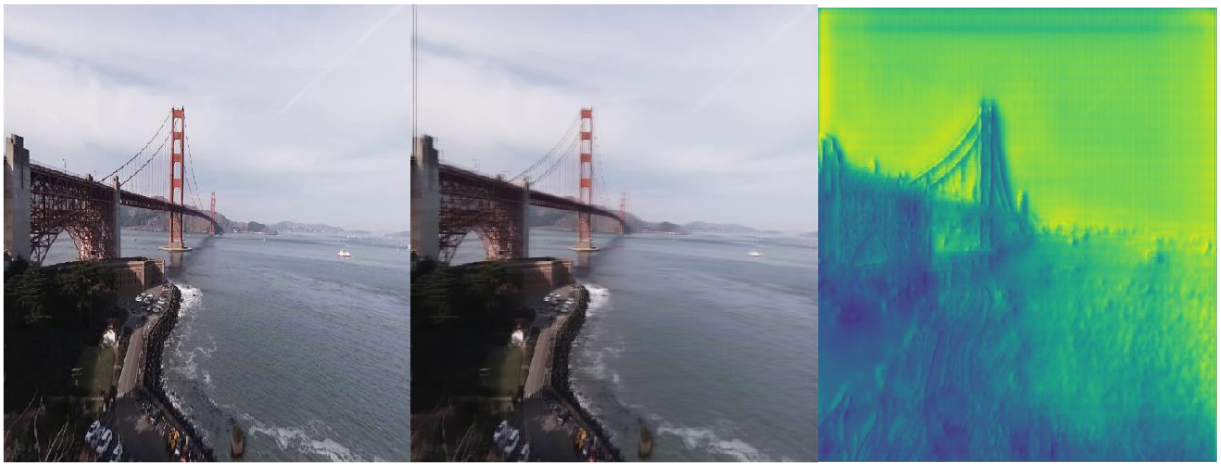
Pozitīvi piemēri

Šajā apakšnodaļa attēloti piemēri, kur dziļuma datu noteikšana ir nostrādājusi veiksmīgi, un iegūtie attēli ļoti līdzinās īstenībai (Skat. *1.-4. att.*). Kreisajā pusē ir redzams ievadītais attēls, pa vidu ģenerētais attēls, un labajā pusē disparitāšu matrica, kurā tumšāka krāsa nozīmē lielāku nobīdi pa kreisi. Attēlos joprojām ir redzams neliels izplūdums, bet izšķirtspējai palielinoties, tas kļūst mazāk nozīmīgs.

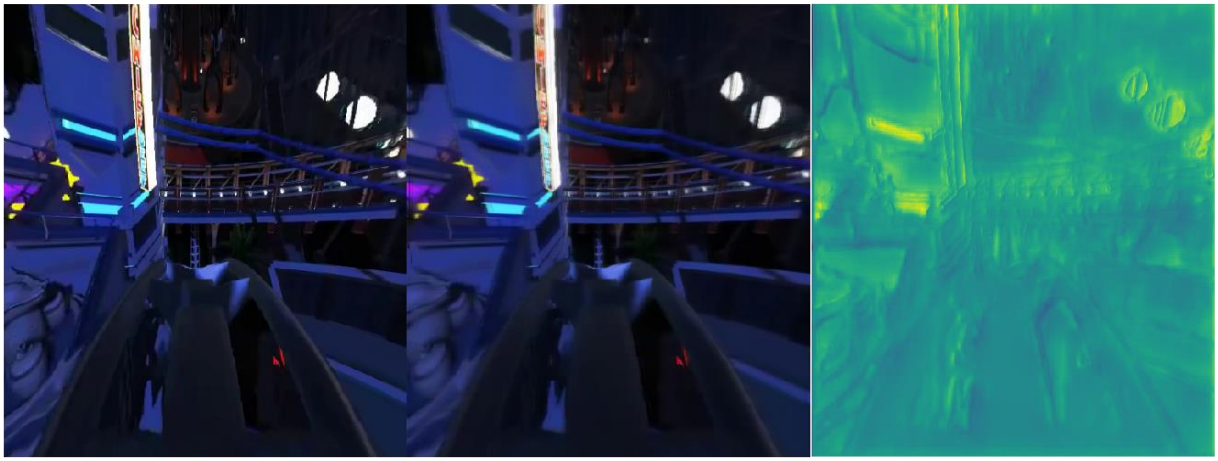
1. att. **Ģenerēta stereoskopiska attēla piemērs**



2. att. **Ģenerēta stereoskopiska attēla piemērs**



3. att. **Ģenerēta stereoskopiska attēla piemērs**



4. att. **Ģenerēta stereoskopiska attēla piemērs**

Negatīvi piemēri

Šajos attēlos ir novērotas kādas būtiskas nepilnības, kuru dēļ tie ir viegli atšķirami no īstiem attēliem. *5. att.* ir iegūts, izmantojot no 3d filmu datu kopas uz VR datu kopu pārtrenētu tīklu, un ir pārāk izplūdis, jo disparitāšu mērogs abās datu kopās netika koriģēts un atšķīrās.

5. att. **Izplūduša ģenerēta attēla paraugs**

6. att. ir virspusēji pieņemams, bet baltais panorāmas rats, kas ir redzams fonā, tiek uzskatīts par baltās telts sastāvdaļu, un, skatoties ar stereoskopiskiem līdzekļiem, pievērš uzmanību tam, ka attēls visdrīzāk ir mākslīgi ģenerēts. Šādas kļūdas varētu mēģināt risināt ar lielāku un specifiskāku datu kopu pielietošanu trenēšanā, kā arī ar laicisko datu izmantošanu, jo tie satur dziļuma informāciju, kas nav atkarīga no paša objekta formas vai apkārtējās vides.

6. att. **Nepareizas dziļuma informācijas paraugs**

SECINĀJUMI

No iegūtajiem rezultātiem un labākā modeļa konfigurācijas var secināt, ka visvarīgākais veiksmīga risinājuma izveidē ar mākslīgajiem neironu tīkliem ir treniņdatu apjoms un daudzveidība. Sarežģītas modeļa struktūras un datu apstrāde var palīdzēt sniegt lielāku efektivitāti, un pārāk vienkāršs modelis var vispār nespēt problēmu atrisināt, bet apjomīga treniņdatu kopa ļaus tīklam attīstīties krietni tālāk un efektīvāk.

Ir pāragri spriest par pētītā risinājuma spēju praktiski un kvalitatīvi ģenerēt VR saturu jebkuros apstākļos, bet no šī darba rezultātiem un izveidotā risinājuma[7] ir skaidrs, ka šāda metode tehniski strādā, un tai ir potenciāls atrisināt pašreizējo VR satura trūkumu. Tādēļ, darba mērķi ir uzskatāmi par izpildītiem.

Hipotēzes

Ar šo metodi ir iespējams iegūt cilvēkam stereoskopiski interpretējamu otras acs attēlu.
Hipotēze apstiprinās. Darba autors personīgi ir apskatījis neskaitāmus stereoskopiskus attēlus, kas ģenerēti ar šo metodi, un telpiskuma izjūta ir acīmredzama.

Atmiņas un trenēšanas laika ierobežojumi neļauj šo tieši izdarīt praktiskā 1920x1080 izšķirtspēja.
Hipotēze neapstiprinās. Ģenerētā materiāla izšķirtspēja nav tieši atkarīga no treniņdatu izšķirtspējas, un darba autors ir uzģenerējis vairākus attēlus 2048x1152 izšķirtspējā, kas pārsniedz prasīto 1920x1080 izšķirtspēju.

Ģenerējot VR video materiālus rezultējošā kadra vidējā absolūtā kļūda būs vismaz divas reizes lielāka, nekā ģenerējot parastus stereoskopiskus kadrus ne-VR 3D filmām.
Hipotēze apstiprinās. Rezultātu nodaļā var redzēt, ka ar vienu un to pašu neironu tīklu, ģenerējot no VR video iegūtus stereoskopiskus kadrus, kļūda ir 0,0581, izmantojot 3D filmu treniņdatus, un 0,0637, izmantojot VR treniņdatus, bet ģenerējot parastus 3D filmu kadrus, tikai 0,0201. Tas ir skaidrojams ar to, ka VR video kadros ir vairāk telpiskas informācijas, kā arī šie kadri aizņem lielāku skata leņķi, tādēļ tie ir sarežģītāk interpretējami un mazāk vienmērīgi.

[7] https://bitbucket.org/karlis383/vrnet/

Iespējamie uzlabojumi un tālākie pētījumi

Ļoti bieži šajā darbā gadījās, ka pēc kādas izmaiņas kļūdas aprēķins rādīja labāku vērtību, bet pārbaudot pašu attēlu, izrādījās, ka tas ir kļuvis krietni sliktāks. Turpretī, kļūdai ceļoties, reizēm notika arī būtiska faktiskās kvalitātes uzlabošanās. Tas ļauj secināt, ka uz pikseļiem bāzēta vidējā absolūtā kļūda nav īsti laba vērtības zuduma funkcija šim risinājumam, un būtu lietderīgi atrast efektīvāku alternatīvu vai papildinājumu.

Mēģinot trenēt tīklu ar ne-VR 3D filmu treniņdatiem, un pēc tam pārtrenējot ar VR datiem, radās nederīgs modelis, jo fokusa plaknes, mērogs un, līdz ar to, arī tīkla spēja paredzēt pikseļu nobīdi krasi mainījās. Lai to nepieļautu un spētu veiksmīgi pārtrenēt tīklu, būtu nepieciešams atrast kādu mērogošanas faktoru un mehānismu, kā šo faktoru iekļaut rezultējošās bildes aprēķinā.

Ne vienmēr visa informācija par doto ainu ir pieejama no vienas bildes, tādēļ būtu lietderīgi izmantot trešo dimensiju video materiālos - laiku. Ar pareizi uzkonstruētu rekurentu neironu tīklu kustīgā video materiālā dziļuma datus varētu noteikt pat neatkarīgi no tā, kas tur attēlots, tikai izmantojot ģeometriju. Turklāt, citi kadri varētu arī dot informāciju par tām nelielajām kadra daļām, kuras sākotnējam skatam ir aizsegtas, un līdz ar to tīklam nav redzamas.

Vēl viena metode, kas arī izmanto laika informāciju, bet to apstrādā jau pirms padošanas pašam tīklam, ir *optical flow*, jeb optiskā kustība. Tā ir informācija par to, kādā virzienā dotajā brīdī kusās katrs pikselis. Ja papildus parastam kadram tīklam padod arī šādu informāciju, pēc perspektīvas likumiem pozīciju varētu noteikt krietni precīzāk.

Strādājot ar platleņķa attēlu datiem, nākas saskarties ar projekcijām, kropļojumiem un sarežģītiem aprēķiniem starp dažādām koordinātām. Iespējams, ka no tā var izvairīties, risinājumā izmantojot sfēriskus konvolūciju tīklus. [37]

Un visbeidzot, lai šo risinājumu vai jebkuru no tā uzlabojumiem varētu pielietot praktiski, ir daudz precīzāk jāizpēta konkrētās metodes kvalitātes iezīmes un robežas, un to, cik labi un patstāvīgi šī metode ir spējīga sniegt rezultātus arī praktiskos apstākļos.

IZMANTOTĀS LITERATŪRAS SARAKSTS

1. David Matthews *Virtual-reality applications give science a new dimension* Nature, 557, 127-128 (2018)
2. *What is Virtual Reality?* (2017) [tiešaite] vrs.org.uk Pieejams: https://www.vrs.org.uk/virtual-reality/what-is-virtual-reality.html [Pārbaudīts 26.05.2018]
3. *Virtual reality production: Where do I start?* (2017) [tiešaite] bbc.co.uk Pieejams: http://www.bbc.co.uk/guides/z3m7k2p#zym7k2p [Pārbaudīts 26.05.2018]
4. David Cardinal *Titanic: How do you convert a movie to 3D, anyway?* (2012) [tiešsaite] extremetech.com Pieejams: https://www.extremetech.com/extreme/124965-titanic-how-do-you-convert-a-movie-to-3d-anyway
5. John D. Pettigrew *The evolution of binocular vision* [tiešsaite] uq.edu.au Pieejams: http://www.uq.edu.au/nuq/jack/BinocVisEvol.pdf [Pārbaudīts 26.05.2018]
6. Jann Paul Mattern, Katja Fennel and Michael Dowd *Introduction and Assessment of Measures for Quantitative Model-Data Comparison Using Satellite Images* (2010) [tieešsaite] mdpi.com Pieejams: http://www.mdpi.com/2072-4292/2/3/794
7. *Depth Map from Stereo Images* [tiešsaite] opencv.org Pieejams: https://docs.opencv.org/3.0-beta/doc/py_tutorials/py_calib3d/py_depthmap/py_depthmap.html [Pārbaudīts 26.05.2018]
8. B. Baykant Alagoz *Obtaining Depth Maps From Color Images By Region Based Stereo Matching Algorithms* (2008) [tiešsaite] arxiv.org Pieejams: https://arxiv.org/ftp/arxiv/papers/0812/0812.1340.pdf [Pārbaudīts 26.05.2018]
9. Nikos Paragios, Yunmei Chen, Olivier D. Faugeras *Handbook of Mathematical Models in Computer Vision* ASV, Springer Science & Business Media, 2006
10. Vilayanur S. Ramachandran, Diane Rogers-Ramachandran *Two Eyes, Two Views: Your Brain and Depth Perception* (2009) [tiešaite] scientificamerican.com Pieejams: https://www.scientificamerican.com/article/two-eyes-two-views/ [Pārbaudīts 26.05.2018]
11. Djilali Amrani *Malus's law of light polarization using a Computer-Based Laboratory* (2009) [tiešsaite] researchgate.net Pieejams: https://www.researchgate.net/figure/Light-intensity-I-as-a-function-of-the-angle-th_fig1_26628775 [Pārbaudīts 26.05.2018]

12. Abdulaziz Alshaer Holger Regenbrecht David O'Hare *Immersion factors affecting perception and behaviour in a virtual reality power wheelchair simulator* (2016) [tiešsaite] sciencedirect.com Pieejams: https://www.sciencedirect.com/science/article/pii/S0003687016300904 [Pārbaudīts 26.05.2018]
13. Gordon Wetzstein *Head Mounted Display Optics I* [tiešsaite] stanford.edu Pieejams: https://stanford.edu/class/ee267/lectures/lecture7.pdf
14. *What is Motion-To-Photon Latency?* (2015) [tiešsaite] chioka.in Pieejams: http://www.chioka.in/what-is-motion-to-photon-latency/ [Pārbaudīts 26.05.2018]
15. http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA295861[Pārbaudīts 26.05.2018]
16. *Overview of Positional Tracking Technologies for Virtual Reality* (2014) [tiešsaite] roadtovr.com Pieejams: https://www.roadtovr.com/overview-of-positional-tracking-technologies-virtual-reality/ [Pārbaudīts 26.05.2018]
17. Noam Kroll *28mm Lenses: The Secret Ingredient For Achieving A Film Look* (2013) [tiešsaite] noamkroll.com Pieejams: http://noamkroll.com/28mm-lenses-the-secret-ingredient-for-achieving-a-film-look/ [Pārbaudīts 26.05.2018]
18. Sensor Size & Field of View (2015) [tiešsaite] panavision.com Pieejams: https://www.panavision.com/sites/default/files/docs/documentLibrary/2%20Sensor%20Size%20FOV%20%283%29.pdf [Pārbaudīts 26.05.2018]
19. Raul Rojas *Neural Networks: A Systematic Introduction* Springer Science & Business Media, 2013 [Pārbaudīts 26.05.2018]
20. Zhang, Wei *Parallel distributed processing model with local space-invariant interconnections and its optical architecture. (1990).* [tiešsaite] drive.google.com Pieejams: https://drive.google.com/file/d/0B65v6Wo67Tk5ODRzZmhSR29VeDg/view?usp=sharing [Pārbaudīts 26.05.2018]
21. Arden Dertat *Applied Deep Learning - Part 4: Convolutional Neural Networks* (2017) [tiešsaite] towardsdatascience.com Pieejams: https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2 [Pārbaudīts 26.05.2018]
22. Jonathan Huang Vivek Rathod Chen Sun Menglong Zhu Anoop Korattikara Alireza Fathi Ian Fischer Zbigniew Wojna Yang Song Sergio Guadarrama Kevin Murphy *Speed/accuracy trade-offs for modern convolutional object detectors* (2017) [tiešsaite] thecvf.com Pieejams: http://openaccess.thecvf.com/content_cvpr_2017/papers/Huang_SpeedAccuracy_Trade-Offs_for_CVPR_2017_paper.pdf [Pārbaudīts 26.05.2018]

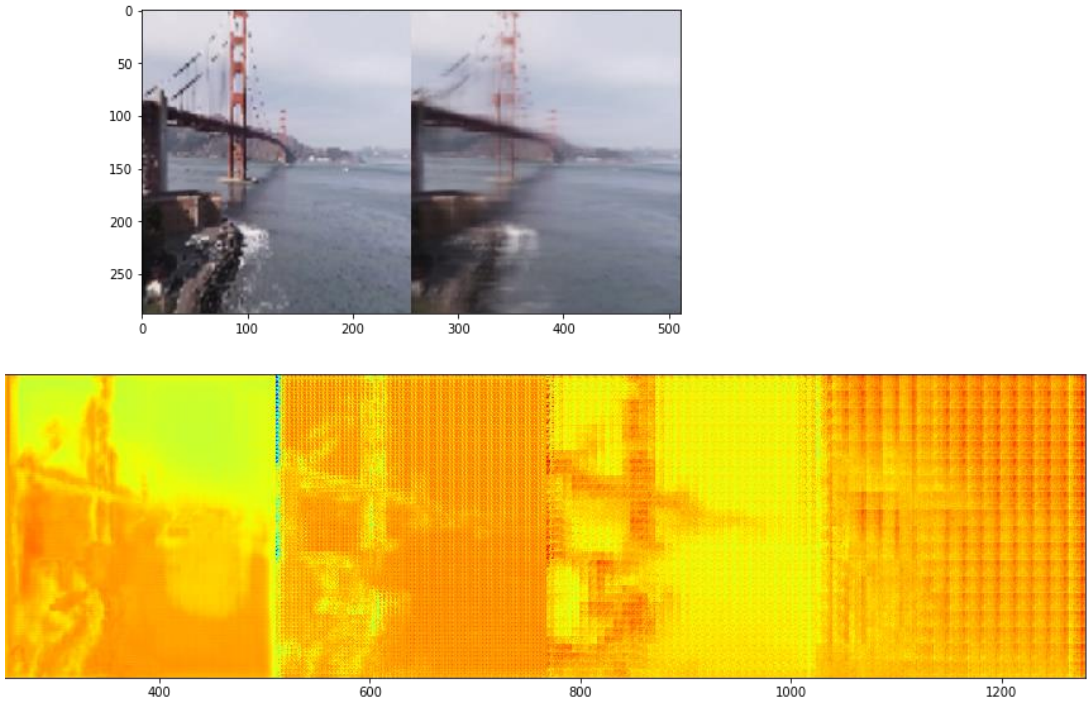
23. Deep Learning | NVIDIA Developer [tiešsaite] nvidia.com Pieejams: https://developer.nvidia.com/deep-learning [Pārbaudīts 26.05.2018]
24. Brando Amos *Image Completion with Deep Learning in TensorFlow* (2018) [tiešsaite] bamos.github.io Pieejams: http://bamos.github.io/2016/08/09/deep-completion/ [Pārbaudīts 26.05.2018]
25. Emin Orhan, Xaq Pitkow *Skip Connections Eliminate Singularities* (2018) [tiešsaite] openreview.com https://openreview.net/forum?id=HkwBEMWCZ [Pārbaudīts 08.05.2018]
26. Francois Chollet *Xception: Deep Learning with Depthwise Separable Convolutions* (2016) [tiešsaite] arxiv.org Pieejams: https://arxiv.org/pdf/1610.02357.pdf [Pārbaudīts 26.05.2018]
27. Junyuan Xie, Ross Girshick, Ali Farhadi *Deep3D: Fully Automatic 2D-to-3D Video Conversion with Deep Convolutional Neural Networks* (2016) [tiešsaite] arxiv.org Pieejams: https://arxiv.org/pdf/1604.03650.pdf [Pārbaudīts 26.05.2018]
28. Lous Foucard *W-net: Self-Supervised Learning of Depthmap from Stereo Images* (2017) [tiešsaite] github.com Pieejams: https://github.com/LouisFoucard/w-net [Pārbaudīts 26.05.2018]
29. Limin Wang, Yuanjun Xiong, Zhe Wang, Yu Qiao *Towards Good Practices for Very Deep Two-Stream ConvNets* (2015) [tiešsaite] arxiv.org Pieejams: https://arxiv.org/abs/1507.02159 [Pārbaudīts 26.05.2018]
30. Equidistant Cylindrical (Plate Carrée) [tiešsaite] proj4.org Pieejams: https://proj4.org/projections/eqc.html [Pārbaudīts 26.05.2018]
31. Nitish S. Mutha *How to map Equirectangular projection to Rectilinear projection* (2017) [tiešsaite] http://blog.nitishmutha.com Pieejams: blog.nitishmutha.com/equirectangular/360degree/2017/06/12/How-to-project-Equirectangular-image-to-rectilinear-view.html [Pārbaudīts 26.05.2018]
32. *How to correctly adjust your lens spacing on the HTC Vive* (20016) [tiešsaite] vrheads.com Pieejams: https://www.vrheads.com/how-correctly-adjust-lens-spacing-htc-vive [Pārbaudīts 26.05.2018]
33. *VR180* [tiešsaite] vr.google.com Pieejams: https://vr.google.com/vr180/ [Pārbaudīts 26.05.2018]
34. *Structural Similarity Index (SSIM) for measuring image quality* [tiešsaite] mathworks.com Pieejams: https://www.mathworks.com/help/images/ref/ssim.html [Pārbaudīts 26.05.2018]
35. *Camera Calibration and 3D Reconstruction* [tiešsaite] opencv.com Pieejams: https://docs.opencv.org/2.4/modules/calib3d/doc/camera_calibration_and_3d_reconstruction.html#void%20stereoRectify(InputArray%20cameraMatrix1,%20InputArray%2

0distCoeffs1,%20InputArray%20cameraMatrix2,%20InputArray%20distCoeffs2,%20Size%20imageSize,%20InputArray%20R,%20InputArray%20T,%20OutputArray%20R1,%20OutputArray%20R2,%20OutputArray%20P1,%20OutputArray%20P2,%20OutputArray%20Q,%20int%20flags,%20double%20alpha,%20Size%20newImageSize,%20Rect*%20validPixROI1,%20Rect*%20validPixROI2) [Pārbaudīts 08.05.2018]

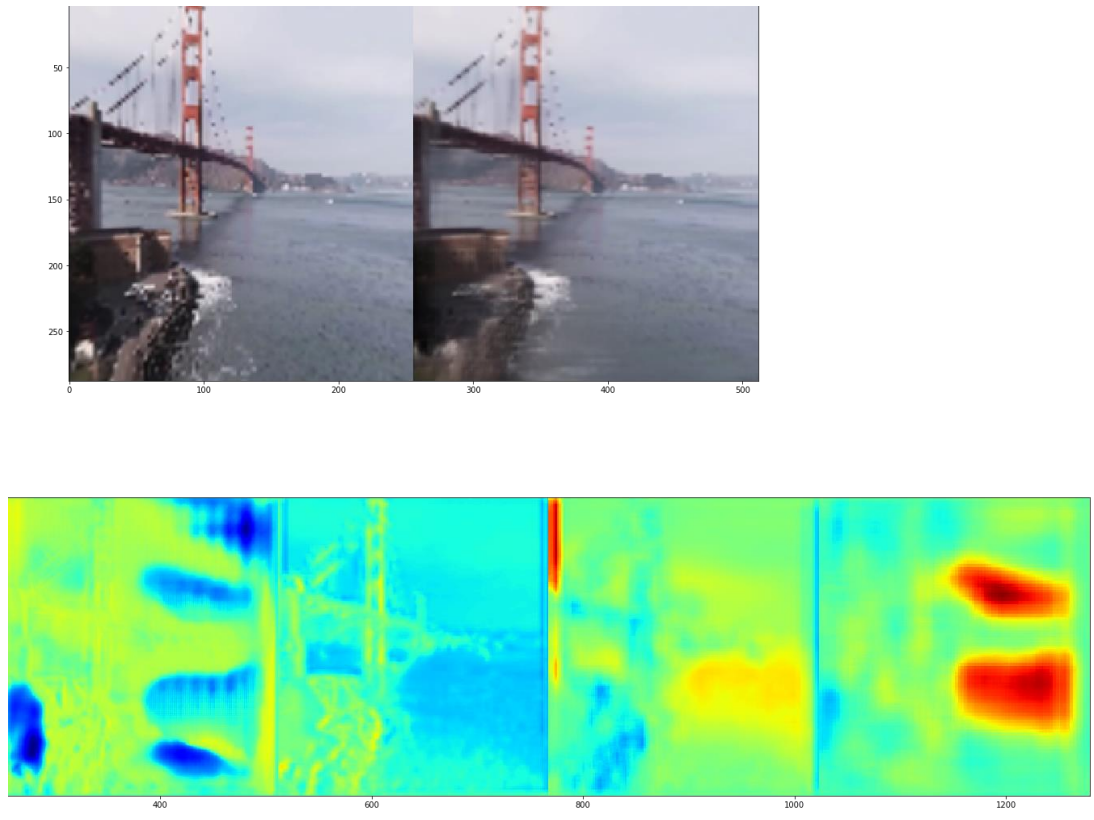
36. *Diederik P. Kingma, Jimmy Ba* Adam: A Method for Stochastic Optimization (2014) [tiešsaite] arxiv.org Pieejams: https://arxiv.org/abs/1412.6980 [Pārbaudīts 08.05.2018]
37. Taco S. Cohen, Mario Geiger, Jonas Koehler, Max Welling *Spehrical CNNs* (2018) [tieešsaite] arxiv.org Pieejams: https://arxiv.org/pdf/1801.10130.pdf [Pārbaudīts 26.05.2018]

PIELIKUMI

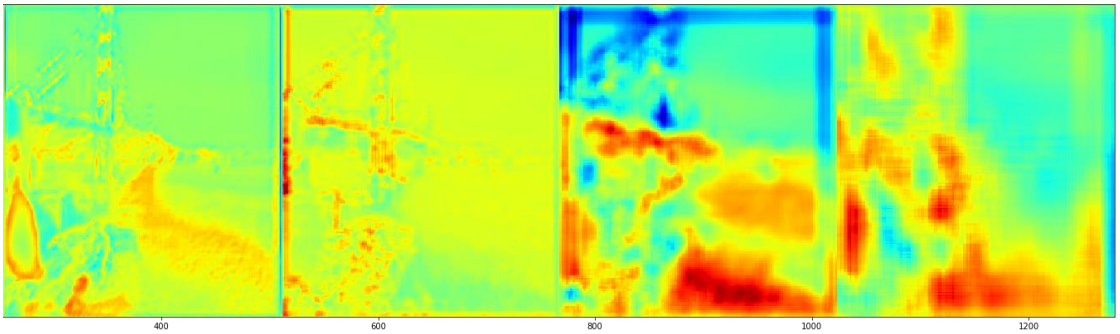
1. Disparitāšu slāņu apvienojuma salīdzinājums dažādiem modeļiem



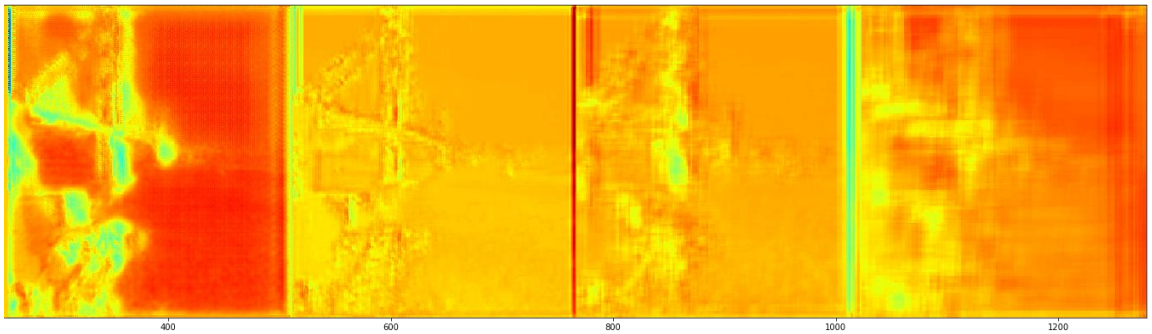
Disparitātes slāņu apvienošana ar transponētu konvolūciju un vidējo aritmētisko



Disparitātes slāņu apvienošana ar transponētu konvolūciju, 1x1 konvolūciju un slāņa dziļumu, kurš ir vienāds ar disparitātes kartes dziļumu



Disparitātes slāņu apvienošana ar transponētu konvolūciju, 1x1 konvolūciju un slāņa dziļumu, kurš ir divas reizes mazāks par disparitātes kartes dziļumu



Disparitātes slāņu apvienošana ar 1x1 konvolūciju, apkopojošo slāni un izkliedētu konvolūciju

DOKUMENTĀRĀ LAPA

Bakalaura darbs „ Stereoskopiska video materiāla ģenerācija no parastiem video materiāliem, izmantojot mašīnmācīšanos” izstrādāts Latvijas Universitātes Datorikas fakultātē.

Ar savu parakstu apliecinu, ka darbs izstrādāts patstāvīgi, izmantoti tikai tajā norādītie informācijas avoti un iesniegtā darba elektroniskā kopija atbilst izdrukai.

Autors: *Kārlis Grošs* _____ 28.05.2018.

Rekomendēju darbu aizstāvēšanai
Darba vadītājs: *Dr. dat. Pēteris Paikens* _____ 28.05.2018**.**

Recenzents: zinātniskais asistents, Mg.sc.comp. Ilvars Mizniks

Darbs iesniegts Datorikas fakultātē 28.05.2018**.**
Dekāna pilnvarotā persona:
vecākā metodiķe Ārija Sproģe_____

Darbs aizstāvēts bakalaura darbu pārbaudījuma komisijas sēdē
___.06.2018. prot. Nr. _____

Komisijas sekretārs(-e): _____