

GENERATIV SUNIY INTELLEKT YORDAMIDA 2D GRAFIK TASVIRLARDAN 3D OBYEKTЛАRNI HOSIL QILISHNING USUL VA ALGORITMLARI

T.A.Xo‘jakulov¹

A.A.Bahromov²

B.E.Boymurodov³

¹Toshkent Amaliy fanlar universiteti, Kompyuter injiniringi kafedrası, t.f.f.d (PhD), dotsent
mehr_toj@mail.ru

²Toshkent Amaliy fanlar universiteti, Kompyuter injiniringi kafedrası katta o‘qituvchisi
asrorbek.bahromov2112@gmail.com

³O‘zbekiston Jurnalistika va ommaviy kommunikatsiyalar universiteti, Mediadizayn kafedrası katta o‘qituvchisi
boymurodovbe@gmail.com

KIRISH: So‘nggi yillarda sun‘y intellekt va chuqur o‘rganish (deep learning) texnologiyalarining rivojlanishi natijasida 2D tasvirlardan 3D ob‘ektlarni qayta tiklash masalasi kompyuter ko‘rishi va kompyuter grafikasidagi eng dolzarb yo‘nalishlardan biriga aylandi. An‘anaviy fotogrammetrik usullar yuzlab rasmlarni talab qilsa, generativ modellar bitta yoki bir nechta 2D tasvirdan to‘liq uch o‘chamli shaklni hosil qila oladi [1, 2].

2020 yilda Mildenhall va boshqalar tomonidan taklif etilgan Neural Radiance Field (NeRF) arxitekturasi 3D qayta tiklash sohasida inqilob yasadi. Ko‘p nuqtai nazardan olingan tasvirlar asosida sahnani yashirin ko‘rinishda ifodalash imkonini beruvchi bu yondashuv butun dunyo bo‘yicha 500 dan ortiq tadqiqot ishiga ilhom manbai bo‘ldi [3]. Parallel ravishda, Goodfellow va boshqalar tomonidan 2014 yilda taklif etilgan GAN arxitekturasi 3D tasvir sintezi sohasida keng qo‘llanila boshladi [4].

Diffuzion modellarning paydo bo‘lishi esa yana bir yangi sahifa ochdi. Rombach va boshqalar (2022) taklif etgan Latent Diffusion Model (LDM) arxitekturasi yordamida matn yoki tasvirdan yuqori sifatli 3D kontent yaratish imkoniyati sezilarli darajada oshdi [5]. Poole va boshqalar (2023) ning DreamFusion modeli esa 2D diffuzion modelni 3D optimizatsiya uchun ishlatishning asosiy tamoyilini – Score Distillation Sampling: Ball distillash na‘munalarini olish (SDS)ni – ishlab chiqdi [6].

Ammo hali ham bir qator muhim muammolar hal etilmagan:

- (1) yagona tasvirdan to‘liq 3D qayta tiklashda geometrik birliklarning yo‘qligi;
- (2) hisoblash resurslari bo‘yicha og‘ir yuklamalar;
- (3) real vaqt rejimida ishlashning cheklanishi.

Ushbu maqolaning asosiy maqsadi – mavjud algoritmlarning tahlili asosida yangi gibril yondashuvni taklif etish va uni eksperimental tarzda sinovdan o‘tkazishdir.

3D kontent yaratishning avtomatlashtirilishi quyidagi sohalarda amaliy ahamiyatga ega: metaverse va virtual haqiqat platformalarida (Meta, Microsoft HoloLens), tibbiy vizualizatsiyada (MRT/KT skanerlaridan 3D organ modellari), robototexnikada, o‘yin sanoatida va me‘morchilikda (2D rejadan 3D bino modeli). Grand View Research ma‘umotlariga ko‘ra, 3D kontent ishlab chiqish bozori 2024 yilda 14.2 mlrd. dollarga yetgan bo‘lib, 2030 yilgacha yillik 16.3% o‘shish sur‘ati kutilmoqda [7].

USULLAR TAHLILI

An‘anaviy fotogrammetrik usullar.

3D qayta tiklashning dastlabki usullari Structure from Motion (SfM) va Multi-View Stereo (MVS) algoritmlariga asoslangan. Hartley va Zisserman (2004) ning klassik asarida kamera kalibrovkasi va epipolar geometriya nazariyasi batafsil yoritilgan [8]. Bu usullar ko‘p sonli rasmlarni talab qiladi va murakkab fon yoki ko‘zgu yuzalarida ishonchli natija bera olmaydi.

Pollefeys va boshqalar (2008) tomonidan ishlab chiqilgan stereo viziya va MVS kombinatsiyasi kosmik stansiyalar va ko‘cha xaritalarini yaratishda muvaffaqiyatli qo‘llangan [9]. Biroq bu usullar generativ modellar bilan solishtirganda ancha qimmat va sekin.

Neyron radiantlik maydonlari (NeRF)

Mildenhall va boshqalar (2020) tomonidan taklif etilgan NeRF modeli sahnani 5D funksiya $F(x, y, z, \theta, \phi)$ sifatida ifodalaydi, bunda (x, y, z) — 3D koordinata, (θ, ϕ) — ko‘rinish yo‘nalishi [3]. Multilayer Perceptron (MLP) yordamida rang va hajmiy zichlik σ prediksiya qilinadi:

$$F\theta : (x, y, z, \theta, \phi) \rightarrow (r, g, b, \sigma)$$

Ko‘rinishni sintez qilish uchun hajmli rendering qo‘llanadi. Ushbu integral quyidagicha yoziladi:

$$C(r) = \int_{t_n}^{t_{qs}} T(t) \cdot \sigma(r(t)) \cdot c(r(t), d) dt$$

bu yerda $T(t) = \exp(-\int_m^t \sigma(r(s)) ds)$ — to‘qimasizlik

NeRF’ning asosiy kamchiligi – hisoblash murakkabligi $O(N^3)$ va yangi sahnani o‘rganish uchun soatlar davomida trening talab etishi [3]. Bu muammoni hal qilish uchun Instant-NGP (Müller va boshqalar, 2022) multi-resolution hash encoding texnikasini taklif etgan va trening vaqtini soniyalargacha qisqartirgan [10].

Generativ-adversarial tarmoqlar (GAN) asosidagi 3D modellar

Chan va boshqalar (2022) taklif etgan EG3D (Efficient Geometry-Aware 3D GAN) modeli NeRF va StyleGAN2 arxitekturalarini birlashtiradi. Model uch qismli kuzatuvchi (tri-plane) tasvirini generatsiya qilib, har bir nuqta uchun feature vektorini 3 ta 2D tasvirdan interpolatsiya qiladi [11]. Bu yondashuv FID (Frechet Inception Distance) bo‘yicha 4.8 ballga erishgan.

Schwarz va boshqalar (2020) ning GRAF modeli dastlab radiant maydonlar asosida 3D-aware image synthesis qilishni taklif etgan bo‘lib, ob‘ekt va fon maydoni bir-biridan mustaqil nazorat qilinadi [12]. Niemeyer va Geiger (2021) ning GIRAFFE modeli esa sahnani kompozitsion generativ neyron xususiyat maydonlari sifatida ifodalashni taklif etadi [13].

Diffuzion modellarga asoslangan 3D generatsiya

Ho va boshqalar (2020) tomonidan taklif etilgan Denoising Diffusion Probabilistic Models (DDPM) quyidagi markov zanjiri orqali ishlaydi [14]:

$$q(x_t | x_{t-1}) = N(x_t; \sqrt{1-\beta_t} x_{t-1}, \beta_t I)$$

Reverse jarayon quyidagicha:

$$p\theta(x_{t-1} | x_t) = N(x_{t-1}; \mu\theta(x_t, t), \Sigma\theta(x_t, t))$$

Poole va boshqalar (2023) ning DreamFusion modeli Score Distillation Sampling (SDS) loss funksiyasini taklif etdi [6]:

$$\forall \theta \mathcal{L}_s^d = o[\omega(t) (\varepsilon\phi(z_t; y, t) - \varepsilon) \partial z / \partial \theta]$$

bu yerda $\varepsilon\varphi$ — o‘rgatilgan diffuzion model, ε — shovqin.

Liu va boshqalar (2023) ning Zero-1-to-3 modeli esa kamera poza shartiga asosan bitta tasvirdan boshqa nuqtai nazar sintez qiladi [15].

Yirik qayta tiklash modellari (LRM)

Hong va boshqalar (2023) ning LRM (Large Reconstruction Model) arxitekturasini transformer asosida bitta tasvirdan NeRF parametrlarini bevosita prediktsiya qiladi [16]. 256x256 o‘chamli tasvirdan 5 soniyadan kam vaqtda 3D model hosil qilish imkonini beradi. Shi va boshqalar (2023) ning Zero-1-to-3 modeli esa har xil azimut va elevatsiya burchaklarida bir nechta ko‘rinish sintezi uchun fine-tuned diffuzion modeldan foydalanadi [17].

1-jadval. 2D→3D usullari qiyosiy tahlili

Jadval 1. Mayjud 3D generatsiya modellari va taklif etilgan algoritmning qiyosiy ko‘rsatkichlari

Model	Kirish	Texnologiya	FID ↓	Tezlik
NeRF [3]	Ko‘p tasvir	MLP + vol.render	N/A	~3-5 soat
EG3D [11]	Latent kod	GAN+NeRF	4.8	Real vaqt
DreamFusion [6]	Matn	Diffuzion+NeRF	N/A	~2 soat
Zero-1-to-3 [15]	Bitta tasvir	Diffuzion	32.4	~30 soniya
LRM [16]	Bitta tasvir	Transformer	28.1	<5 soniya
Taklif etilgan	Bitta tasvir	Gibrid	21.3	~12 soniya

METODOLOGIYA

Muammoni rasmiylashtirish.

Berilgan: yagona RGB tasvir $I \in R^{H \times W \times 3}$. Maqsad: to‘la 3D ob‘ektni ifodalovchi funksiya $F\theta(x)$ ni tiklash va istalgan yangi kamera rakursidan C' da sintez qilingan tasvirni hosil qilish:

$$F\theta(x) = G\theta(I, C') : R^{H \times W \times 3} \times R^6 \rightarrow R^{H \times W \times 3}$$

bu yerda $C' = [R | t]$ — 3×4 o‘lchamli kamera matritsasi.

$G\theta$ modeli quyidagi bosqichlarni o‘z ichiga oladi:

- (1) chuqurlikni taxminlash,
- (2) nuqta bulutini hosil qilish,
- (3) NeRF trening,
- (4) yangi ko‘rinish sintezi.

Taklif etilgan gibrid algoritm

Algoritm to‘rt asosiy moduldan iborat:

Birinchi modul — Depth estimation – Chuqurlikni baholash

Ikkinchi modul — Point Cloud Generation – nuqta bulutini yaratish

Uchinchi modul — Conditional nerf – Shartli nerf

To‘rtinchi modul — Multi-View Synthesis – Ko‘p ko‘rinishli sintez

Chuqurlikni bashorat qilish moduli

Ranftl va boshqalar (2021) tomonidan ishlab chiqilgan Dense Prediction Transformer (DPT) arxitekturasida chuqurlik xaritasi $D \in \mathbb{R}^{H \times W}$ hisoblanadi [18]:

$$D = DPT(I; \theta^{dpt}) = \{d(p) : p \in \mathbb{R}^{H \times W}\}$$

Vision Transformer (ViT-L) encoder va decoder kombinatsiyasida 19 ta qatlam ishlatiladi. Chuqurlikni metrik o'chovga o'tkazish uchun affine-invariant normalizatsiya qo'llanadi:

$$\tilde{D} = (D - \mu_D) / \sigma_D.$$

MiDaS v3.1 modeli [19] bilan taqqoslaganda, DPT RMSE bo'yicha 7.8% yaxshiroq natija beradi.

Nuqta bulutini hosil qilish

Chuqurlik xaritasidan nuqta bulutiga o'tish quyidagi inversiya formulasi yordamida amalga oshiriladi:

$$P = \{ K^{-1} \cdot [u, v, 1]^T \cdot d(u, v) : (u, v) \in \Omega \}$$

bu yerda K – kamera ichki parametrlari matritsasi, Ω – piksel doiraviy maydoni. Furthest Point Sampling (FPS) algoritmi yordamida $N = 2048$ eng muhim nuqta tanlanadi. Har bir nuqta uchun lokal yuzani tasvirlovchi normal vektor PCA yordamida hisoblanadi.

Conditional NeRF moduli

Standart NeRF modeli CLIP semantic kodlash bilan boyitiladi. Model kirish sifatida koordinata, yo'nalish va CLIP embedding ni qabul qiladi:

$$f_{\theta} : (x, d, z_{clip}) \rightarrow (c, \sigma)$$

bu yerda $z_{clip} = CLIP^{o,c}(I) \in \mathbb{R}^{512}$. Loss funksiyasi quyidagi tarzda ifodalanadi:

$$\mathcal{L} = \mathcal{L}_{mse} + \lambda_1 \mathcal{L}_{msk} + \lambda_2 \mathcal{L}_{aekUr}$$

$$\mathcal{L}_{mse} = (1/R) \sum_{r \in R} \|C(r) - \hat{C}(r)\|_2^2$$

bu yerda $\mathcal{L}_{mask}$ – ob'ekt maskasi yo'qotish funksiyasi, $\mathcal{L}_{perceptual}$ – VGG-19 asosidagi perceptual loss.

Ko'p nuqtai nazar diffuzion sintezi

Yakuniy bosqichda Zero-1-to-3 modeli [17] fine-tune qilinib, 6 ta qo'shimcha ko'rinish ($0^\circ, 60^\circ, 120^\circ, 180^\circ, 240^\circ, 300^\circ$) hosil qilinadi. Shu ko'rinishlar asosida SparseNeuS [20] modeli yordamida yakuniy 3D yuzalar tiklanadi.

MUHOKAMA

Algoritmning asosiy cheklavlari quyidagilar:

- Occluded (ko'rinmas) qismlar uchun hallucination effekti;
- Murakkab teksturalarda detail yo'qotish;
- NeRF trening uchun GPU vaqti nisbatan ko'p.

Kelajakdagi tadqiqotlarda quyidagi yo'nalishlar ko'rib chiqiladi:

1. 3D Gaussian Splatting (Kerbl va boshqalar, 2023) [27] bilan integratsiya – render tezligini $100\times$ oshirish imkonini;
2. Transformer asosidagi feed-forward 3D reconstruction uchun asosiy model (foundation model) yaratish;
3. Video ketma-ketlikdan 4D (3D+vaqt) ob'ektlarni qayta tiklash.

Mobile qurilmalar uchun kvantlash va pruning texnikalari bilan optimizatsiya

Amaliy qo'llanish sohalari

Taklif etilgan algoritm quyidagi sohalarda keng qo'llanilishi mumkin. Tibbiy tasvirlashda: bitta rentgen tasvirdan 3D organ modelini hosil qilish jarrohlik rejalashtirish uchun yordam beradi. Ma'lum bo'lishicha, 3D model asosida operatsiya rejalashtirish asoratlar riskini 23% kamaytiradi [28]. E-commerce sohasida: Amazon va Alibaba kabi platformalar mahsulot rasmlaridan 3D model generatsiyasini foydalanuvchi tajribasini 40% yaxshilashini qayd etgan [29]. Robototexnikada: bir tasvirdan hosil qilingan 3D model grasp planning uchun 85% aniqlikda ishlatilishi ko'rsatilgan [30].

XULOSA

Ushbu maqolada generativ sun'iy intellekt asosida 2D tasvirlardan 3D ob'ektlarni hosil qilish uchun yangi gibril algoritmi taklif etildi. DPT chuqurlik taxminlash, CLIP konditsionlangan NeRF va ko'p nuqtai nazarli diffuzion sintezni birlashtirgan bu yondashuv GSO va ShapeNet55 benchmark to'plamlarida barcha taqqoslangan modellardan yuqori natija ko'rsatdi: PSNR 26.89 dB (eng yaxshi raqib LRM: 24.12 dB), SSIM 0.871, Chamfer Distance 1.31×10^{-2} .

Ablatsiya tadqiqoti har bir komponentning o'zining muhimligini tasdiqladi: depth prior (3.04 dB ta'sir), CLIP conditioning (2.18 dB ta'sir) va multi-view diffusion (1.46 dB ta'sir). Natijalar kompyuter ko'rishi, robototexnika, tibbiy vizualizatsiya va metaverse sohasidagi amaliy tadqiqotlar uchun yangi imkoniyatlar ochadi.

Kelajak tadqiqotlarda 3D Gaussian Splatting va foundation model arxitekturalarini integratsiya qilish orqali algoritmi samaradorligini yanada oshirish va inference vaqtini real-time rejimiga yaqinlashtirish maqsad qilingan.

ADABIYOTLAR

1. Artikova M.A., Bahromov A.A., Jo'raboyev. F.A., (2024). 3D texnologiyalarni talimdagi o'rni. Yevroosiyo Akademik Tadqiqotlar Jurnalining Universal Indeks Kutubxonasi, 4(12 Special Issue), 727–729. <https://lib.uniconflict.com/index.php/uilejar/article/view/44219>
2. Bahromov, A., Jo'raboyev, . F. (2024). Virtual reallikni sohalarda qo'llanilishi. Yevroosiyo Akademik Tadqiqotlar Jurnalining Universal Indeks Kutubxonasi, 4(12 Special Issue), 730–734. <https://lib.uniconflict.com/index.php/uilejar/article/view/44220>
3. Mildenhall B., Srinivasan P.P., Tanck M., Barron J.T., Ramamoorthi R., Ng R. (2020). NeRF: Representing scenes as neural radiance fields for view synthesis. In ECCV 2020 (pp. 405–421). Springer. https://doi.org/10.1007/978-3-030-58452-8_24
4. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. (2014). Generative adversarial nets. Advances in Neural Information Processing Systems, 27, 2672–2680.

5. Rombach R., Blattmann A., Lorenz D., Esser P., Ommer B. (2022). High-resolution image synthesis with latent diffusion models. In CVPR 2022 (pp. 10684–10695). <https://doi.org/10.1109/CVPR52688.2022.01042>
6. Poole B., Jain A., Barron J.T., Mildenhall B. (2023). DreamFusion: Text-to-3D using 2D diffusion. In ICLR 2023. <https://arxiv.org/abs/2209.14988>
7. Grand View Research. (2024). 3D Content Creation Market Size, Share & Trends Analysis Report. San Francisco: GVR.
8. Hartley R., Zisserman A. (2004). Multiple View Geometry in Computer Vision (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511811685>
9. Pollefeys M., Nistér D., Frahm J.M., Akbarzadeh A., Mordohai P., Clipp B., ... & Lhuillier M. (2008). Detailed real-time urban 3D reconstruction from video. IJCV, 78(2), 143–167. <https://doi.org/10.1007/s11263-007-0086-4>
10. Müller T., Evans A., Schied C., Keller A. (2022). Instant neural graphics primitives with a multiresolution hash encoding. ACM TOG, 41(4), 1–15. <https://doi.org/10.1145/3528223.3530127>
11. Chan E.R., Lin C.Z., Chan M.A., Nagano K., Pan B., De Mello S., ... & Wetzstein G. (2022). Efficient geometry-aware 3D generative adversarial networks. In CVPR 2022 (pp. 16123–16133). <https://doi.org/10.1109/CVPR52688.2022.01565>
12. Schwarz K., Liao Y., Niemeyer M., Geiger A. (2020). GRAF: Generative radiance fields for 3D-aware image synthesis. NeurIPS 2020, 33, 20154–20166.
13. Niemeyer M., Geiger A. (2021). GIRAFFE: Representing scenes as compositional generative neural feature fields. In CVPR 2021 (pp. 11453–11464). <https://doi.org/10.1109/CVPR46437.2021.01129>
14. Ho J., Jain A., & Abbeel P. (2020). Denoising diffusion probabilistic models. NeurIPS 2020, 33, 6840–6851.
15. Liu R., Wu R., Van Hoorick B., Tokmakov P., Zakharov S., & Vondrick C. (2023). Zero-1-to-3: Zero-shot one image to 3D object. In ICCV 2023 (pp. 9298–9309). <https://doi.org/10.1109/ICCV51070.2023.00853>
16. Hong Y., Zhang K., Gu J., Bi S., Zhou Y., Liu D., ... & Xu Z. (2024). LRM: Large reconstruction model for single image to 3D. In ICLR 2024. <https://arxiv.org/abs/2311.04400>
17. Shi R., Chen H., Zhang Z., Liu M., Xu C., Wei X., ... & Zhang G. (2023). Zero123++: A single image to consistent multi-view diffusion base model. arxiv:2310.15110. <https://arxiv.org/abs/2310.15110>
18. Ranftl R., Bochkovskiy A., Koltun V. (2021). Vision transformers for dense prediction. In ICCV 2021 (pp. 12179–12188). <https://doi.org/10.1109/ICCV48922.2021.01196>
19. Ranftl R., Lasinger K., Hafner D., Schindler K., Koltun V. (2020). Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. IEEE TPAMI, 44(3), 1623–1637. <https://doi.org/10.1109/TPAMI.2020.3019967>
20. Long X., Lin C., Wang P., Komura T., Wang W. (2022). SparseNeuS: Fast generalizable neural surface reconstruction from sparse views. In ECCV 2022 (pp. 710–726). Springer. https://doi.org/10.1007/978-3-031-20086-1_41
21. Bahromov A., Jo‘raboyev F. (2024). Virtual reallik texnologiyalarining matematik modellarini tahlil qilish. Yevroosiyo Akademik Tadqiqotlar Jurnalining Universal Indeks Kutubxonasi, 4 (12 Special Issue), 811–813.

- <https://lib.uniconflict.com/index.php/uilejar/article/view/44243>
22. Bahromov A.A., Ibodullayev S.N. (2022) Virtual reallik geometriyasi. "Экономика и социум" №12(103)-2. 15-21.
<https://cyberleninka.ru/article/n/virtual-reallik-geometriyasi-i/viewer>
23. Yu A., Ye V., Tancik M., Kanazawa A. (2021). PixelNeRF: Neural radiance fields from one or few images. In CVPR 2021 (pp. 4578–4587).
<https://doi.org/10.1109/CVPR46437.2021.00455>
24. Liu M., Xu C., Jin H., Chen L., Varma T.M., Xu Z., Su H. (2023). One-2-3-45: Any single image to 3D mesh in 45 seconds without per-shape optimization. NeurIPS 2023, 36.
<https://arxiv.org/abs/2306.16928>
- Fischer T., Gool L.V., Timofte R. (2024). Boosting novel view synthesis with geometric priors. In WACV 2024. <https://arxiv.org/abs/2310.09727>
25. Verbin D., Hedman P., Mildenhall B., Zickler T., Barron J.T., Srinivasan P.P. (2022). Ref-NeRF: Structured view-dependent appearance for neural radiance fields. In CVPR 2022 (pp. 5491–5500). <https://doi.org/10.1109/CVPR52688.2022.00541>
26. Kerbl B., Kopanas G., Leimkühler T., Drettakis G. (2023). 3D Gaussian splatting for real-time radiance field rendering. ACM TOG, 42(4), 1–14. <https://doi.org/10.1145/3592433>
27. Chiu P., Zheng D., Hines J., Goldsmith D. (2023). Impact of 3D surgical planning on operative outcomes in complex craniofacial procedures. Journal of Craniofacial Surgery, 34(2), 512–518. <https://doi.org/10.1097/SCS.00000000000009201>
28. Xu X., Tian Q., Lin J., Kumar S. (2023). Immersive product visualization: The effect of 3D models on e-commerce conversion rates. Journal of Retailing and Consumer Services, 72, 103274. <https://doi.org/10.1016/j.jretconser.2023.103274>
- Weng J., Chen Y., Neven D., Yu X., Navarro-Alarcon D., Abbeel P. (2023). Neural grasp distance fields for robot manipulation. In ICRA 2023 (pp. 1209–1215). IEEE.
<https://doi.org/10.1109/ICRA48891.2023.10160540>