

Določanje slikovnega prostora na umetniških slikah

Reconstruction of image space depicted on artistic paintings

Nadezhda Komarova
Gregor Anželj
nadezhdakomarova7@gmail.com
gregor.anzelj@gmail.com
Gimnazija Bežigrad
1000 Ljubljana, Slovenia

Borut Batagelj
Narvika Bovcon
Franc Solina
borut.batagelj@fri.uni-lj.si
narvika.bovcon@fri.uni-lj.si
franc.solina@fri.uni-lj.si
Fakulteta za računalništvo in informatiko
Univerza v Ljubljani

POVZETEK

V članku poročamo o analizi slikovnega prostora na umetniških slikah s pomočjo metod računalniškega vida. Naš cilj je bil, da ugotovimo, ali je možno zgojiti na osnovi zaznave obrazov na slikah določiti prostorsko organizacijo slike. Analiza je potekala na izbranem vzorcu 3356 slik. Najprej smo določili tridimenzionalne koordinate zaznanih obrazov na posamezni sliki. Nato smo tem točkam priredili ravnino. Slikovni prostor smo tako določili z enačbo prirejene ravnine oziroma kotom med to ravnino in slikovno ravnino. Bolj kot je ravnina, ki jo določajo obrazi, nagnjena od navpične smeri, globlji je prikazani slikovni prostor.

KLJUČNE BESEDE

računalniški vid, slikovni prostor, zaznava obrazov, umetnostna zgodovina

ABSTRACT

In the article, we report on the analysis of the image space depicted on artistic paintings utilizing methods of computer vision. Our aim was to find out whether one can recover the spatial organization of a picture based on detection of faces. The analysis was conducted on the sample of 3356 paintings. First, 3D coordinates of faces were determined. Then, a plane was fitted to the faces on every painting. Images were therefore described in terms of the angle between the fitted plane and the picture plane. The bigger the angle between both planes, the deeper the picture space depicted.

KEYWORDS

computer vision, image space, face detection, art history

1 UVOD IN MOTIVACIJA

Odločili smo se povezati dve raziskovalni področji, ki sta si navidez zelo vsaksebi, to je umetnostna zgodovina in umetna inteligenca. Metode računalniškega vida se že redno uporabljajo tudi za analizo umetniških slik [12]. Večina teh raziskav je osredotočena na analizo posameznih ali manjšega števila umetniških slik. Po drugi strani smo danes v dobi velepodatkov (angl. *Big Data*), saj je vedno več informacij dostopnih v digitalni obliki. Tudi velike zbirke reprodukcij umetniških slik so danes prosto

dostopne na medmrežju, na primer Google Arts and Culture, Wikimedia Commons, Getty Open Content Program, ADA (Archive of Digital Art) in druge [4]. Z analizo in vizualizacijo velikih umetniških zbirk se je prvi začel ukvarjati Lev Manovich [8]. Leta 2012 je preučeval vizualizacijske metode za družboslovne vede in medijske raziskave. Ukvarjal se je z informativno, uporabno in estetsko vrednostjo vizualizacij [9].

Analiza razlik med predstavitvijo prostora s fotografijo in umetniško sliko je bila narejena leta 2014 [11]. S statistično analizo slik tihožitij, ki so jih ustvarili udeleženci eksperimenta, so ugotovili, da so predmeti, na katere so udeleženci usmerjali pozornost, naslikani večji kot so na fotografijah. Zato je vprašanje, ali je dosledna uporaba linearne perspektive najbolj primerna metoda za posnemanje sveta [1]. Umetnostna zgodovina nam nazorno prikaže, da so umetniki za posnemanje sveta uporabljali zelo različne pristope.

Pri naši analizi slikovnega prostora smo izhajali iz dveh predpostavk:

- (1) v raziskavi želimo analizirati veliko število umetniških slik v smislu današnjega trenda *Big Data*,
- (2) uporabiti želimo take metode računalniškega vida, ki delujejo hitro in čimbolj zanesljivo.

Med hitre in zanesljive metode računalniškega vida zagotovo sodi zaznava in identifikacija oseb na osnovi njihovih obrazov. Zaradi varnostnih razlogov se je teh problemov na področju biometrije lotilo že zelo veliko znanstvenikov. Danes obstajajo hitre in zanesljive metode za zaznavo in analizo obrazov na slikah [10].

Za navdih nam je služil članek Irvinga Zupnicka iz leta 1959 [14], objavljen še veliko pred uporabo računalnikov v likovni umetnosti, ki opisuje kako je na slikah iz različnih umetnostnih obdobjih organiziran slikovni prostor. Zato smo si zastavili vprašanje, ali je mogoče s pomočjo metod računalniškega vida rekonstruirati slikovni prostor na umetniških slikah? Bolj konkretno, ali ga je mogoče rekonstruirati na osnovi zaznave obrazov na slikah? Določitev 3D razsežnosti prostora, upodobljenega na sliki, smo se lotili na osnovi pozicije obrazov na sliki (x in y koordinate) in njihove velikosti, kar nam daje grobo informacijo o tretji dimenziji z – to je oddaljenosti obraza od opazovalca. Ta pristop seveda temelji na predpostavki, da so na slikah ljudje oziroma da so upodobljeni njihovi dovolj veliki obrazi. Resda v zgodovini likovne umetnosti poznamo veliko tihožitij ali pokrajinskih slik, na katerih ni obrazov. Toda velika večina umetniških slik iz obdobja pred izumom fotografije dejansko upodablja ljudi oz. njihove obraze.

Iz javno dostopnih baz umetniških slik smo za našo študijo izbrali testno množico 3356 slik iz različnih umetnostnozgodovinskih obdobji in žanrov.

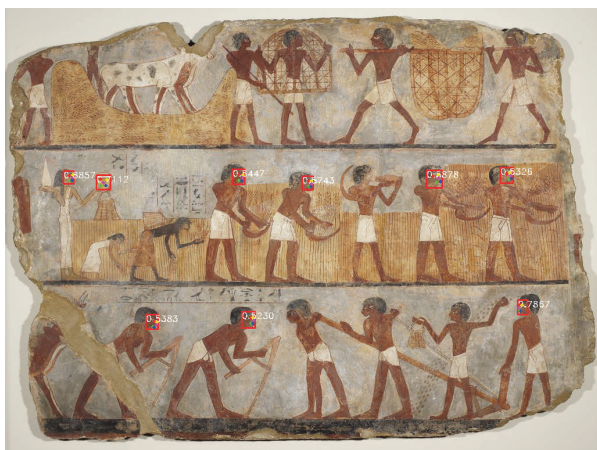
Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Information Society 2021, 4–8 October 2021, Ljubljana, Slovenia
© 2021 Copyright held by the owner/author(s).

2 SLIKOVNI PROSTOR NA UMETNIŠKIH SLIKAH



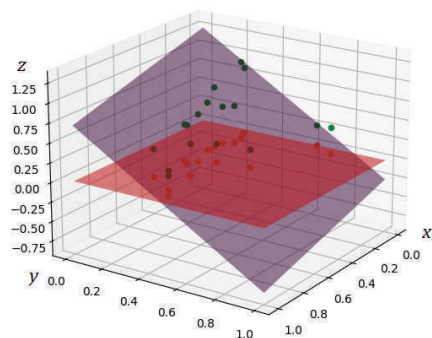
Slika 1: Auguste Renoir, *Ples v Le Moulin de la Galette*; vidijo se zaznani obrazi. Velikost obrazov jasno odraža globino slikarskega prostora.



Slika 2: Poslikava v grobnici *Unsu*. Vsi obrazi se enake velikosti, ves slikarski prostor je zgoščen kar v ravnini poslikave.

Vsakemu obrazu na slikah smo privedili tridimenzionalne koordinate, ki pa niso bile zanesljive v absolutnih vrednostih, temveč odražajo zgolj relativne razdalje. Nato smo tem obraznim točkam privedili ravnine s smislu vsote najmanjših kvadratov razdalj med točkami in iskano ravnino. Pri tistih slikah, ki prikazujejo obraze, ki so v spodnjem delu slike opazovalcu blizu, in se višje na sliki postopno oddaljujejo (Slika 1), so dobljene ravnine bolj nagnjene v globino kot pri tistih, kjer so vsi obrazi približno na enaki razdalji od opazovalca (Slika 2). V takih primerih je dobljena ravnina skorajda vzporedna s površino slike. Rafaelova *Atenska šola* in staroegipčanska poslikava v grobnici *Unsu* imata zelo različni prostorski ureditvi. Na prvi sliki se obrazi zmanjšujejo z oddaljevanjem ljudi. Ravnina, prirejena točkam na tej sliki, je zato nagnjena v globino (Slika 3).

Po drugi strani tudi poslikava na Sliki 2 prikazuje množico ljudi, vendar so vsi enake višine in njihovi obrazi so enako veliki. Ravnina, prirejena obrazom na egipčanski sliki, je zato vzporedna ravnini $z = 0$. Za egipčansko slikarstvo je značilno konceptualno



Slika 3: Vijoličasta ravnina, ki se prilega 3D pozicijam obrazov na Renoirjevem *Plesu v Le Moulin de la Galette* in in rdeča ravnina $z = 0$ – ploskev slikarskega platna, na kateri smo zaznali obraze.

upodabljanje prostora: velikosti oseb niso določene s prostorskim oddaljevanjem, temveč npr. z družbenim statusom.

3 ZAZNAVA OBRAZOV

Predpostavili smo, da so resnični obrazi pri vseh osebah približno enako veliki. Zato so bili večji obrazi obravnavani kot bližji površini slike in manjši kot bolj oddaljeni od površine slike oz. od opazovalca.

Zaznani so bili z orodjem *RetinaFace*, ki izvede dvodimenzionalno poravnano in tridimenzionalno rekonstrukcijo obraza [2]. Zasnovan je na osnovi globoke nevronske mreže.

Detektor vrne podatke o obrazih v dvodimenzionalnem prostoru površine slike, torej imajo središča obraznih okvirjev in točke oči, nosu ter ust samo x in y koordinate. Toda za rekonstrukcijo tridimenzionalnega prostora slike potrebujemo tudi globine obrazov oz. koordinato z . Tridimenzionalni prostor, kot ga prikazuje umetniška slika, se razlikuje od fotografskega predvsem zato, ker slikarji redko dosledno upoštevajo linearno perspektivo. Na fotografijah je perspektiva po drugi strani bolj konsistentno določena. Zato je na njih mogoče z enačbo (1) [6] določiti oddaljenost predmeta od kamere:

$$d = \frac{f \cdot h_r \cdot h}{h_i \cdot h_s} \quad (1)$$

Z enačbo (1) izračunamo oddaljenost d objekta v milimetrih, če je f goriščna razdalja fotoaparata, h_r resnična višina objekta v milimetrih, h višina slike v pikslih, h_i višina objekta na sliki v pikslih in h_s višina sensorja fotoaparata v milimetrih. Z njo so bile določene tudi oddaljenosti obrazov na slikah v vzorcu, pri čemer so bile uporabljene vrednosti goriščne razdalje in višine sensorja, kvocient katerih opiše, kako vidijo človeške oči. Četudi je bilo po tem postopku nemogoče določiti natančne tridimenzionalne koordinate obrazov na sliki, so bile določene relativne oddaljenosti med obrazi in površino slike. Za namen te raziskave tudi niti ni pomembno, če zaznamo vse obraze na sliki.

4 GEOMETRIJSKA INTERPRETACIJA PROSTORA

Parametre A , B in C enačbe ravnine $z = Ax + By + C$ smo določili z minimizacijo funkcije

$$E(A, B, C) = \sum_{i=1}^m (Ax_i + By_i + C - z_i)^2, \quad (2)$$

kjer m pomeni število točk in x_i , y_i ter z_i koordinate točk. Funkcija (2) doseže minimum, ko je $\nabla E = (0, 0, 0)$ [3]. Za gradient te funkcije velja $\nabla E = (\frac{\partial E}{\partial A}, \frac{\partial E}{\partial B}, \frac{\partial E}{\partial C})$, kjer so $\frac{\partial E}{\partial A}$, $\frac{\partial E}{\partial B}$ in $\frac{\partial E}{\partial C}$ naslednji.

$$\frac{\partial E}{\partial A} = 2 \sum_{i=1}^m x_i (Ax_i + By_i + C - z_i) \quad (3)$$

$$\frac{\partial E}{\partial B} = 2 \sum_{i=1}^m y_i (Ax_i + By_i + C - z_i) \quad (4)$$

$$\frac{\partial E}{\partial C} = 2 \sum_{i=1}^m (Ax_i + By_i + C - z_i) \quad (5)$$

Tako množici 3D točk priredimo ravnino z minimizacijo razdalj med temi točkami in njihovimi slikami na ploskvi v smeri z . Koeficienti A , B in C so zato rešitve sistema linearnih enačb (6), (7) in (8).

$$A \sum_{i=1}^m x_i^2 + B \sum_{i=1}^m x_i y_i + C \sum_{i=1}^m x_i = \sum_{i=1}^m x_i z_i \quad (6)$$

$$A \sum_{i=1}^m x_i y_i + B \sum_{i=1}^m y_i^2 + C \sum_{i=1}^m y_i = \sum_{i=1}^m y_i z_i \quad (7)$$

$$A \sum_{i=1}^m x_i + B \sum_{i=1}^m y_i + C = \sum_{i=1}^m z_i \quad (8)$$

5 REZULTATI

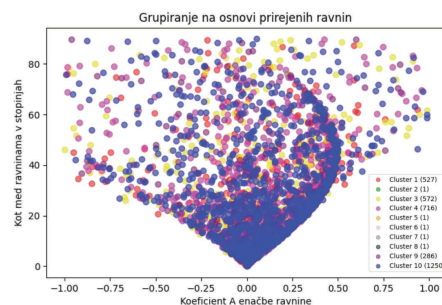
Slike smo izbrali iz prostodostopne zbirke WikiArt (<https://www.wikiart.org>), kjer so umetnine med drugim razdeljene po žanrih. Izbrana so bila slikarska dela (potrebno je bilo izločiti npr. kiparska), kjer je bilo upodobljenih več ljudi. Iz zbirke WikiArt so bila zato izbrana dela iz žanrov *pastorale* (77 slik), *allegorical painting* (1225 slik), *history painting* (1377 slik) in *literary painting* (667 slik), in sicer skupaj 3356 slik. Poleg žanra smo imeli tudi podatke o umetnostno zgodovinskem obdobju v katero sodi posamezna slika. Zanimalo nas je, kako lahko le na osnovi teh podatkov smiselno razdelimo testno množico slik z metodo gručenja in ali je ta delitev relevantna z vidika umetnostno zgodovine.

Kot kriterij pri gručenju so bile uporabljene enačbe ravnin ter kot med prirejen ravnino in slikovno ravnino $z = 0$. Detektor *RetinaFace* opiše slednje s tremi parametri – rotacijami okoli osi x , y in z (v pozitivni in negativni smeri). Pri posamezni sliki so bile izbrane rotacije v vsaki smeri z največjimi absolutnimi vrednostmi.

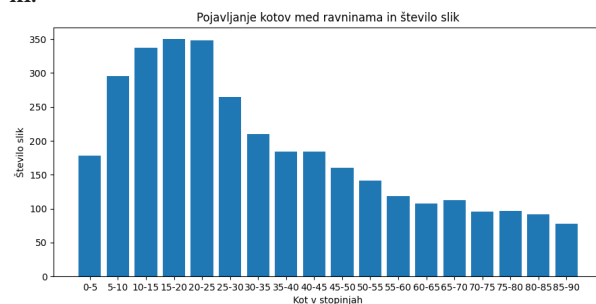
Gručenje je bilo opravljeno z algoritmom BIRCH, implementiranim s knjižnico *scikit-learn*. BIRCH (angl. *Balanced Iterative Reducing and Clustering using Hierarchies*) je algoritem gručenja, ki je posebej prilagojen delu z večjimi podatkovnimi vzorci [7].

Na Sliki 4 so ekstremne vrednosti izločene. Prikazana je razporeditev slik po gručenju na osnovi ravnin. Bila je izvedena primerjava tega, katerim umetnostnim slogom pripadajo slike v posameznih razredih. To je bilo mogoče, saj je bila vsaka slika v zbirki označena poleg žanra tudi z letom nastanka in umetnostnim slogom (barok, romantika ipd.). Število razredov smo omejili na deset. Zaradi izrazite drugačnosti prostorske razporeditve na nekaterih slikah so bile slednje izločene v posamezne razrede (2, 5, 6, 7 in 8). Ti razredi vsebujejo le po eno sliko in niso vidni na Sliki 4.

Histogram na Sliki 5 prikazuje zastopanost različnih intervalov kotov v proučevanem vzorcu. Vidi se, da je bil največji delež slik takih, kjer je bil kot med ravninama med 15 in 20 stopinj, kar se zdi relativno malo. Večji koti med ravninama večinoma



Slika 4: Razporeditev razredov pri gručenju na osnovi ravnin. Gruče so razpršene in izrazite razmejitve med njimi.



Slika 5: Zastopanost posameznih kotov za slike v testni množici.

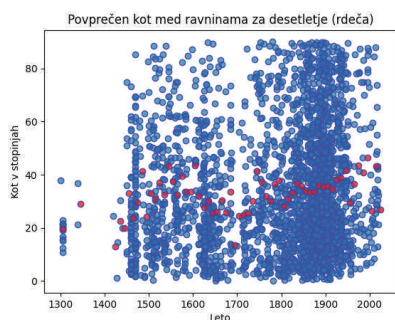
ustrezajo slikam, kjer se upodobljene osebe enotno oddaljujejo oz. približujejo. Če je bil kot med ravnino, ki je bila prirejena obrazom na sliki, in ravnino $z = 0$ izračunan kot natančno 0 stopinj, je to pomenilo, da na sliki ni bilo zaznanega nobenega obraza, samo en obraz ali pa so imeli vsi obrazi enake globine. Na intervalu od 0 do 5 stopinj (Slika 5) je bil najpogostejši barok, na preostalih intervalih po romantika. Ni pa na nobenem intervalu močno prevladoval le en slog, saj je odstotek slik, ki je pripadal najpogostejšemu slogu v posameznem intervalu med 20 in 30%.

Za določitev korelacije med časom nastanka posamezne slike in kotom med ravninama za to sliko je bil uporabljen Spearmanov koeficient korelacije. Ta predstavlja neparametersko stopnjo povezanosti med spremenljivkama oz. kako dobro je mogoče opisati njun odnos z monotono funkcijo [13]. Koeficient je bil 0.183, kar predstavlja šibko pozitivno korelacijo. p vrednost je bila v tem primeru blizu 0, kar pomeni, da korelacija med letom nastanka slike in kotom, ki odraža slikarsko globino ni linearna. Na prikazu na Sliki 6 je razvidno, da če opazujemo obdobje od približno leta 1700 in vse do danes, povprečen kot med ravninama za posamezna desetletja blago narašča.

6 RAZPRAVA

Glavna hipoteza naše raziskave je bila, ali lahko na nek enostaven način ugotovimo kakšen je slikarski prostor, to je, kako izrazita je globinska dimenzija na dani umetniški sliki. Slikarski prostor pa je povezan tako z umetnostno zgodovinskim obdobjem v katerem sodi slika, kot tudi z žanrom slike. Na ta način se nam odpira možnost avtomatske klasifikacije velikega števila slik, bodisi s statističnimi metodami, še bolj pa bi prišle v poštev metode strojnega učenja.

Odločili smo se, da bomo slikovni prostor določali posredno s pomočjo zaznave obrazov. Ko je bil posamezen obraz zaznan z orodjem *RetinaFace*, je bil s tem določen obrazni okvir na določeni



Slika 6: Koti med ravninama v odvisnosti od časa nastanka slike. Rdeče točke predstavljajo povprečen kot za posamezno desetletje.

koordinati x in y na ravnini slike. Velikost obraznega okvirja pa nam je dal še informacijo o relativni oddaljenosti obraza z od ravnine slike. Zanesljivost zaznave obrazov na umetniških slikah je bil verjetno nekoliko slabši, saj je bil *RetinaFace* naučen na fotografijah obrazov in ne na umetniških upodobitvah [2]. V kakšni prihodnji raziskavi bi lahko uporabili še dodatne informacije, ki jih daje orodje *RetinaFace* za zaznavo obrazov: orientacija obraza, lega oči, nosu in ust, spol ter starost osebe, določeno na osnovi obraza. Poleg tega bi lahko v prihodnjih raziskavah pri analizi slik upoštevali tudi barvno sestavo in druge slikovne značilke, ki jih lahko robustno določimo z metodami računalniškega vida [12]. Sami smo se ukvarjali npr. z detekcijo črt perspektivne projekcije na fotografijah [5, 1].

Četudi smo v našem preizkusu metode likovna dela združevali v razrede po podobnosti prostorske ureditve, se niso pokazale stroge meje med umetnostnimi slogi slik. Informativna pa je bila korelacija med časom nastanka dela in kotom med ravninama. V izbranem vzorcu slik različni umetnostnozgodovinski slogi niso bili povsem enakomerno zastopani in je bilo npr. veliko del iz romantike. Za vsako zgodovinsko obdobje so najverjetneje izrazite določene medsebojne povezanosti teh značilnosti. Ustavljen umetnostnozgodovinski pristop pri analizi slik je sočasno opazovanje dveh ali več del, pri katerih raziskovalec na osnovi svojega predhodnega znanja izloči značilne poteze, razlike ipd. [8]. Strojno učenje bi na tej točki postalo učinkovito, saj po eni strani nudi možnost analize velike količine podatkov, odkrivanje sočasnih povezav med različnimi značilkami, po drugi strani pa zagotavlja objektivnost matematičnih pristopov. Zato bi bilo v nadaljevanju koristno uporabiti poleg obrazov tudi druge informacije na slikah. Potrebno pa je upoštevati, da delitev umetniških del ne more biti absolutna, saj umetnostno zgodovino sestavljajo posamezni umetniki, vsak od njih ustvarja v svojem lastnem slogu, ki lahko do neke mere sledi splošnim trendom obdobja, vendar nikoli popolnoma. Tudi posamezni likovni umetniki v času svoje kariere lahko spremenijo svoj umetniški slog.

7 ZAKLJUČEK

V članku smo pokazali nov pristop k avtomatski analizi umetniških slik z uporabo metod računalniškega vida. Demonstrirali smo, da je z metodo zaznave obrazov na slikah možno nasloviti tudi bolj kompleksna vprašanja, kot v našem primeru organizacija prostora na slikah. Čeprav rezultati te raziskave morda niso tako jasno izraženi in niso reproducirali rezultatov umetnostnih zgodovinarjev, se uporaba računalnikov na področju umetnostne zgodovine kot na sploh v humanistiki šele zares začne. Računalniško zasnovane analitične metode bodo omogočile odgovore

na vprašanja, ki si jih umetnostni zgodovinarji do sedaj sploh še niso upali zastaviti.

LITERATURA

- [1] Katarina Bebar. "Upodabljanje prostora po načelih linearne perspektive s pomočjo obogatene resničnosti". V: *Likovne besede* 114 (2020), str. 14–21.
- [2] Jiankang Deng in sod. "RetinaFace: Single-Shot Multi-Level Face Localisation in the Wild". V: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020, str. 5202–5211. DOI: 10.1109/CVPR42600.2020.00525.
- [3] David Eberly. "Least Squares Fitting of Data". V: *Magic Software, Inc.* (sep. 2001). URL: http://www.sci.utah.edu/~balling/FEtools/doc_files/LeastSquaresFitting.pdf.
- [4] *Image resources: Free image resources*. Sotheby's Institute of Art. URL: <https://sia.libguides.com/images/freeimagereources> (pridobljeno 1. 3. 2021).
- [5] Jure Kovač, Peter Peer in Franc Solina. "Automatic natural and man-made scene differentiation using perspective geometrical properties of the scenes". V: *Proceedings 15th International Conference on Systems, Signals and Image Processing*. Bratislava, 2008, 507–510.
- [6] Yun Liang. "How to measure the real size of an object from an unknown picture?". Jan. 2015. URL: <https://www.researchgate.net/post/How-to-measure-the-real-size-of-an-object-from-an-unknown-picture>.
- [7] Cory Maklin. "BIRCH Clustering Algorithm Example In Python". V: *towards data science* (jul. 2019). URL: <https://towardsdatascience.com/machine-learning-birch-clustering-algorithm-clearly-explained-fb9838cbeed9>.
- [8] Lev Manovich. "Data Science and Digital Art History". V: *International Journal for Digital Art History* 1 (jun. 2015). DOI: 10.11588/dah.2015.1.21631.
- [9] Lev Manovich. *Museum without walls, art history without names: visualization methods for Humanities and Media Studies*. Oxford Handbook Online, 2013. DOI: 10.1093/oxfordhb/9780199757640.013.005.
- [10] Mohd Nayeem. "Exploring Other Face Detection Approaches (Part 1) – RetinaFace". V: *Analytics Vidhya* (jul. 2020). URL: <https://medium.com/analytics-vidhya/exploring-other-face-detection-approaches-part-1-retinaface-9b00f453fd15>.
- [11] Robert Pepperell in Manuela Braunagel. "Do Artists Use Linear Perspective to Depict Visual Space?". V: *Perception* 43 (avg. 2014), 395 – 416. DOI: 10.1068/p7692.
- [12] David G. Stork. "Computer Vision and Computer Graphics Analysis of Paintings and Drawings: An Introduction to the Literature". V: *Computer Analysis of Images and Patterns*. Ur. Xiaoyi Jiang in Nicolai Petkov. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, str. 9–24. DOI: 10.1007/978-3-642-03767-2_2.
- [13] Eric W. Weisstein. "Spearman Rank Correlation Coefficient". V: *MathWorld, a Wolfram Web Resource* (brez datuma). URL: <https://mathworld.wolfram.com/SpearmanRankCorrelationCoefficient.html>.
- [14] Irving L. Zupnick. "Concept of Space and Spatial Organization in Art". V: *The Journal of Aesthetics and Art Criticism* (dec. 1959), str. 215–221. DOI: 10.2307/427268.