

МАПИРАЊЕ ТЕКСТУРЕ ЛИЦА ГЕНЕРИСАНЕ ПАМЕТНИМ ТЕЛЕФОНОМ НА ГЕНЕРИЧКИ МОДЕЛ ЛИЦА

MAPPING FACE TEXTURE GENERATED BY SMARTPHONE ONTO GENERIC FACE 3D MODEL

Милица Лазор, Факултет техничких наука, Нови Сад

Област – РАЧУНАРСКА ГРАФИКА

Кратак садржај – Представљено је једноставно и универзално решење, које задовољава потребе корисника а ниво квалитета одговара уређају на ком се приказује, тј. паметном телефону. Основна идеја је могућност брзог креирања персонализованог анимираног аватара који би се користио на друштвеним мрежама, виртуелним конференцијама и слично. Користе се већ готови, усредњени модели главе, са добром топологијом, као и постављеним скелетом. Персонализација модела врши се помоћу текстуре, те она мора бити довољно уверљива. Текстура се добија употребом камере на мобилном телефону. Мапирање слике на модел врши се препознавањем значајних тачака лица помоћу Dlib библиотеке. Кључни алгоритам у овом раду је тзв. Thin Plate Spline алгоритам.

Кључне речи: Текстура, мапирање, лице, детекција, UV мапе, 3Д

Abstract – A simple and universal solution was presented, which satisfies the needs of users and the level of quality corresponds to the device on which it is displayed, i.e. a smartphone. The basic idea is the ability to quickly create a personalized animated avatar that would be used on social networks, virtual conferences, etc. Ready-made, averaged models are used, with a good topology, as well as a set skeleton. The personalization of the model is done with the help of texture, so it must be convincing enough. The texture was obtained using a camera on a mobile phone. Mapping an image onto a model is done by recognizing facial landmarks using the Dlib library. The key algorithm in this paper is the Thin Plate Spline algorithm.

Keywords: Texture, mapping, face, detection, UV maps, 3D

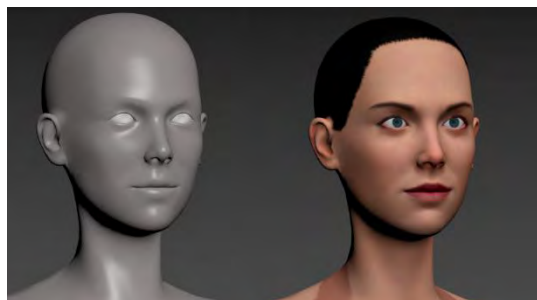
1. УВОД

У последњих пар деценија, направљен је огроман помак у истраживањима у области 3Д реконструкције [1, 2]. Технологија је толико напредовала да је могуће добити скоро реалистичне моделе предмета, ентеријера, екстеријера – ипак, реконструкција људског лица захтева додатан напор [3, 4, 5].

НАПОМЕНА:

Овај рад је проистекао из мастер рада чији ментор је био др Драган Дину, доцент.

И најситније неправилности доводе до тога да такав модел делује неприродно и неуверљиво. Ово се односи како на геометрију, тако и на анимацију, текстуру и све остале сегменте. Креирање једног фотореалистичног модела људског лица поред квалитетног и скупог хардвера захтева и много времена, труда и мануелног рада. Ипак, са појавом и све широм употребом паметних телефона, ствара се потреба за брзим и прорачунски јефтним решењима. Квалитет се не може занемарити, али свакако пада у други план [6]. Значај текстуре јасно се може видети на слици где је приказан генерички 3Д модел без и са текстуром, слика 1.



Слика 1. Приказ истог модела без текстуре и са текстуром

Представљено решење комбинује низ различитих техника из домена компјутерске визије и компјутерске графике. Рад се заснива на проналажењу кореспонденција генеричког 3Д модела и фотографије усликане паметним телефоном. Помоћу Dlib детекције лица и значајних тачака на истом, може се извршити прецизно мапирање текстуре. Велику улогу игра интерполација текстуре између кључних тачака, имплементирана преко Thin plate spline - TPS алгоритма. С обзиром на то да је улазни податак само једна фотографија, један од изазова је прикривање и надомешћивање мањка информација (теме главе, уши, врат). У раду су приказане две методе за решавање овог питања, од којих једна ради по принципу мешања (blending) улазне фотографије са генеричком UV мапом, док се друга бави уклањањем позадине са фотографије и екстраполацијом добијене текстуре.

Као резултат овог рада добија се текстура лица спремна за мапирање на генерички модел главе. Главна предност приказаног решења је брзина и универзалност – безброј различитих фотографија лица могуће је мапирати на исти генерички модел лица, и

тако добити исто толико персонализованих модела. Такође, овај приступ се може применити и на анимираном моделу и тако пружити кориснику анимираног аватара који се може користити на друштвеним мрежама, веб конференцијама, играма, итд. Комбинација приказаних алгоритама даје аутоматско, брзо и робусно решење, уз минималан напор корисника.

2. ПРЕГЛЕД СТАЊА У ОБЛАСТИ

3Д реконструкција је увек актуелна тема за истраживање, а раме уз раме са њом иде и креирање и мапирање текстуре. Неретко се ова два питања решавају заједно, јер су информације добијене из слике неопходне за реконструкцију, сасвим довољне за креирање текстуре [7, 8]. Ипак, и реконструкција и текстурисање из једне слике представљају посебан изазов.

У раду [9] представљен је алгоритам за реконструкцију модела главе на основу једне слике. Систем се састоји из две компоненте, претходни тренинг модела еиген глава и реконструкција 3Д модела главе. Први део заснива се на итеративној шеми поравнавања модела, од грубог до финог (*course-to-fine*), како би се успоставиле кореспонденције између 3Д модела главе у сету података за обуку. Такође, користи се радијална базна функција за успостављање густих кореспонденција између сваког 3Д модела лица и референтног модела, заједно са техником анализе компонентата (*principal component analysis - PCA*) како би се израчунао статистички еиген модел главе. За 3Д реконструкцију лица из једне слике, алгоритам објашњен у раду проналази најбољу линеарну комбинацију база еиген модела, која минимизују енергетску функцију састављену од растојања између одговарајућих тачака лица и усмерене делимичне Хаусдорфове удаљености између контура лица у домену слике. Резултати приказани у раду показују да предложено решење даје добре резултате када је у питању 3Д реконструкција. Иако акценат у овом раду није на текстури, значајан је допринос употреба кључних тачака за проналажење кореспонденција на моделу. Имплементација алгоритма за искривљење слика (*warping*) представљена је у раду [10]. Основна идеја аутора лежи у оптимизацији простора који текстуре покривају. Метода је заснована на равномерном распореду фреквенцијског садржаја на слици. Подручја високе фреквенције се развлаче (*stretch*) док се контракција врши на подручјима ниске фреквенције. Такође, у обзир се узимају и дисторзије које настају приликом мапирања текстуре на модел. Резултујућа текстура може се затим скупити (*shrink*) без губљења детаља у областима високе фреквенције. Важно је напоменути да овај тип оптимизације утиче само на текстуру и њене координате. Ни геометрија модела ни повезаност се не мењају. Искривљење слика постиже се помоћу једноставног вишемрежног релаксационог алгоритма (*multigrid relaxation*).

Приступ чија је циљана употреба управо на паметним уређајима, описан је у раду [6]. Овај метод захтева само једну, по могућности фронталну фотографију

лица, и не захтева јако прецизне кореспонденције између модела и текстуре. Задовољавајуће мапирање постиже се интерактивном шемом, где корисник дефинише кореспонденције преко уређивања кључних тачака и њихових региона утицаја. Мапирање се постиже једноставном линеарном трансформацијом између 3Д модела и текстуре. У овом раду такође се помиње и радијална базна функција, међутим, у време када се приступ развијао (2004. године), паметни уређаји нису били хардверски довољно јаки како би подржали ово решење.

3. МЕТОДОЛОГИЈА

Помоћу *Dlib* библиотеке, врши се детекција битних тачака на лицу на улазној фотографији. С обзиром да се користи претходно креиран генерички 3Д модел, ове битне тачке су на моделу увек на истом месту.

Модел може бити преузет са интернета или претходно креиран. Што се тиче креирања модела, тај посао захтева више мануелног рада, али је довољно једном креирати модел који ће се увек изнова користити као генерички. Први корак је моделовање, са акцентом на топологији. Наредни корак је обележавање битних тачака лица, истих оних које су детектоване на 2Д улазној фотографији. Затим, модел је потребно отпаковати, да би се добила *UV* мапа. Важно је напоменути да *UV* мапа садржи све тачке модела, па тако и претходно обележене битне тачке лица. Такође, модел је потребно риговати и анимирати, како би се добило комплетно решење.

3.1. Аквизиција и припрема података

Решење описано у овом раду, користи као улаз фотографију генерисану паметним телефоном и претходно креиран 3Д модел.

Паметни телефон који је коришћен је *iPhone 11*, са главном камером од 2x12 мегапиксела. Такође, решење је тестирано и са предњом камером, која на овом телефону такође има резолуцију од 12 мегапиксела. Како би се добио жељени квалитет, важно је да корисник приликом фотографисања испуни пар услова:

1. Лице треба да испуњава већину кадра када се слика фронталном камером.
2. Субјекат не би требао да носи наочаре.
3. Коса субјекта би требала да буде склоњена тако да не покрива лице (ако је у питању субјекат са дужом косом).
4. Осветљење би требало да буде неутрално и уједначено.

4. СОФТВЕРСКА ИМПЛЕМЕНТАЦИЈА И РЕЗУЛТАТИ

Имплементација представљеног решења написана је у *python* програмском језику, помоћу *Dlib* и *OpenCV* библиотека. На слици 2, приказани су сви имплементациони блокови, а сваки од њих биће објашњен у даљем тексту.



Слика 2. Приказ имплементационих блокова

Детекција лица на фотографији је први корак у процесу мапирања текстуре на генерички модел. За ове сврхе, коришћен је претходно обучен *Dlib* детектор лица. Овај детектор лица направљен је помоћу класичног хистограма оријентисаних градијената (HOG) у комбинацији са линеарним класификатором (SVM), пирамидом слика, и шемом детекције клизних прозора (*sliding windows*). Ова врста детектора је прилично добра и способна је да открије многе врсте предмета, поред људских лица [25].

Након што је лице детектовано на слици и његова позиција јасно одређена, прелази се на следећи, а уједно и један од важнијих корака у овом решењу – детекција кључних тачака лица (*facial landmarks*). Под кључним тачкама, подразумевају се истакнути региони лица: очи, обрве, нос, уста, брада.. Као и код детекције лица, и овде се користи претходно обучен модел из *Dlib* библиотеке, који представља имплементацију рада из 2014. године написаног од стране Каземија (*Kazemi*) и Саливана (*Sullivan*) [32]. Ова метода се заснива на тренинг скупу означених кључних тачака на лицу. Сlike су мануелно означене и одређују специфичне (x,y) координате региона који окружује сваку структуру лица. Поред слика, метода се ослања и на вероватноћу растојања између парова улазних пиксела. Узимајући ове тренинг податке, група регресионих стабала (*regressing trees*) обучава се тако да процењује позиције кључних тачака директно на основу самих интензитета пиксела. Модел за детекцију кључних тачака лица унутар *Dlib* библиотеке процењује позиције 68 координата (x, y) које обележавају структуре лица. Ове а anotације су део *iBUG 300-W* [33] сета података на којима је *Dlib* модел за предикцију кључних тачака и трениран..

Након што су детектоване кључне тачке на улазној слици, прелази се на процес мапирања. Први корак је тзв. грубо мапирање. На овај начин улазна слика се позиционира тако да се преклапа са *UV* мапом, ипак, никакве веће трансформације унутар саме слике се нису десиле. Ово мапирање постиже се помоћу матрице хомографије.

Прецизно мапирање текстуре представља суштину овог рада. Потребно је претходно нађене тачке на улазној слици и на *UV* мапи, довести до потпуног преклапања. На овај начин, улазна слика се "развлачи" преко *UV* мапе и потребно је урадити интерполацију над међупикселима, како би решење било тачно и визуелно уверљиво. Алгоритам који омогућава овакве манипулације је *Thin plate spline – TPS* [14]. У основи овог алгоритма налази се радијална базна функција - *RBF*. *RBF* представља

кERNELЕ централизоване око контролних тачака x_i . Што је интерполирана тачка x даља од контролне тачке x_i , мања је удаљеност од одговарајућег KERNELА, и самим тиме KERNEL има мањи утицај на ту тачку. Неки KERNELИ, због веће густине околних тачака у близини центра и због величине, могу имати већи утицај на целу контуру. *RBF* за сваку тачку рачуна пондерисану комбинацију удаљености свих KERNELА сумаризовано помоћу функције која следи (1). Ако је дата улазна тачка x , ова функција за глатку интерполацију ће дати одговарајућу искривљену тачку $x'=f(x)$.

$$f(x) = \sum \alpha_i R(x, x_i) \quad (1)$$

Где α_i и представља тежину која одговара радијалној базној функцији $R(x, x_i)$ око тачке x_i . *TPS* алгоритам приказује примену радијалне базне функције на сликама, односно у 2Д простору. Овај алгоритам врши интерполацију површине која пролази кроз сваку контролну тачку. Сет од три тачке генерисаће равну површину. Контролне тачке се могу замислити као ограничења положаја на површини која се савија. Идеална површина је она која се савија најмање и добија се помоћу следеће функције (2).

$$f(x, y) = a_1 + a_2x + a_3y + \sum_{i=1}^n w_i U(|P_i - (x, y)|) \quad (2)$$

Како би се деформисала слика и добило прецизно мапирање, потребно је променити позицију сваког пиксела. Дакле, потребна је информација где ће се пиксел (x, y) са улазне слике А, налазити на слици Б на позицији (x+dx, y+dy). X и Y координате могу се посматрати засебно. Тако се једна површина може користити за интерполацију помераја dx, а друга површина за интерполацију помераја dy. Додатне хипер-површине се могу користити како се повећава димензија слике. Са површинама dx и dy могуће је израчунати померај (dx,dy) у било којој тачки на слици.

Први од два приступа за фино подешавање текстуре састоји се од трансфера боје (*color transfer*) и мешања (*blending*) генеричке *UV* мапе са добијеном текстуром. Други приступ заснива се на уклањању позадине са добијене текстуре и затим екстраполацији текстуре.

Оба приступа имају своје предности и мане. Први приступ даје комплетну *UV* мапу, али се губи на персонализацији модела. Са друге стране, приступ са уклањањем позадине даје уверљиво мапирање што се тиче лица, али остали делови (уши, врат, теме) су покривени комбинацијом боја са улазне текстуре. Ови недостаци се, на срећу, могу сакрити косом. На слици 3, приказана је улазна слика, мапирање првим

приступом, мапирање другим приступом и затим сакривање недостатака косом.



Слика 3. Упоредни приказ резултата

5. ЗАКЉУЧАК

У овом раду, представљен је низ алгоритама и техника који заједно омогућују мапирање било које слике лица направљене паметним телефоном, на генерички модел лица. Приказане су две методе, од којих свака има своје предности, али исто тако захтева и одређену дораду. У складу са применом, резултати нису високог квалитета, али показују велики потенцијал. Приказани модели, уз додатак косе, могу се користити као 3Д аватари, на друштвеним мрежама, онлине конференцијама али и игрицама. Такође, једна од главних предности овог решења је анимирање. Једном анимиран модел, могао би да пружи безброј персонализованих модела. Овај приступ доводи до значајне уштеде времена и ресурса.

Једна од могућности унапређења је свакако увођење ГАН неуронских мрежа и обучавање модела да аутоматски генерише боље UV мапе на основу текстуре. Такође, када би се као улаз користио видео, или више фотографија, проблем који је до сада постојао због недостатка информација био би моментално решен. Исто тако, потребно је смањити генеричност модела, и самим тиме повећати персонализацију. Ово би се могло постићи креирањем базе модела, из које би се на основу улазне фотографије бирао најсличнији. Друго потенцијално решење би била модификација самог модела у складу са улазном текстуром, транслацијом одговарајућих група вертекса.

ЛИТЕРАТУРА

- [1] Po-Han Lee, Jui-Wen Huang, Huei-Yung Lin, 2012, *3D Model reconstruction based on multiple view image capture*, International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS 2012).
- [2] Xian-Feng Han, Hamid Laga, Mohammed Bennamoun, 2019, *Image-based 3D Object Reconstruction: State-of-the-Art and Trends in the*

Deep Learning Era. IEEE Transactions on Pattern Analysis and Machine Intelligence.

- [3] Yasutaka Furukawa, Jean Ponce, 2007 *Accurate, Dense, and Robust Multi-View Stereopsis*, IEEE Transactions on Pattern Analysis and Machine Intelligence.
- [4] Yao Feng, Fan Wu, Xiaohu Shao, Yanfeng Wang, Xi Zhou, 2018, *Joint 3D Face Reconstruction and Dense Alignment with Position Map Regression Network*, ECCV 2018.
- [5] Shu Liang, Linda G. Shapiro, Ira Kemelmacher-Shlizerman., 2016, *Head Reconstruction from Internet Photos*, ECCV 2016.
- [6] Qi Zhao, Jiajun Bu, Mingli Song, 2004, *Towards talking face on portable devices - A flexible and realistic approach*, TENCON 2004.
- [7] J. Riviere, P. Gotardo, D. Bradley, A. Ghosh, T. Beeler, 2020, *Single-Shot High-Quality Facial Geometry and Skin Appearance Capture*, ACM Trans. Graph., Vol. 39, No. 4, Article 81.
- [8] T. Beeler, B. Bickel, P. Beardsley, B. Sumner, M. Gross, 2010, *High-Quality Single-Shot Capture of Facial Geometry*, ACM Transactions on Graphics 29(4).
- [9] Shu-Fan Wang, Shang-Hong Lai, 2006, *Efficient 3D Face Reconstruction from a Single 2D Image by Combining Statistical and Geometrical Information*, Computer Vision - ACCV 2006.
- [10] Laurent Balmelli, Gabriel Taubin, Fausto Bernardin, 2002, *Space-Optimized Texture Maps*, EUROGRAPHICS 2002
- [11] Arun Ponnusamy, 2018, *CNN based face detector from dlib*, Elsevier Science, <https://towardsdatascience.com/cnn-based-face-detector-from-dlib-c3696195e01c> (приступљено, октобар 2020.)
- [12] V.Kazemi, J.Sullivan, 2014, *One millisecond face alignment with an ensemble of regression trees*, 2014 IEEE Conference on Computer Vision and Pattern Recognition
- [13] *Facial point annotations*, <https://ibug.doc.ic.ac.uk/resources/face-point-annotations/>, (приступљено, октобар 2020.)
- [14] Herve Lombaert, *Manual Registration with Thin Plates*, <https://profs.etsmtl.ca/hlombaert/thinplates/> (приступљено, октобар 2020.)

Кратка биографија:



Милица Лазор рођена је у Новом Саду 1995. год. Мастер рад на Факултету техничких наука из области Рачунарске графике – Анимација у инжењерству одбранила је 2020. год.