УДК 621.396.967(075.8)
DOI https://doi.org/10.32782/2663-5941/2022.6/10

Трапезон К.О.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Войналович О.О.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

ОСОБЛИВОСТІ ВИКОРИСТАННЯ ТЕХНОЛОГІЇ PHOTO WAKE-UP ДЛЯ СИСТЕМ МАШИННОГО ЗОРУ

Системи машинного навчання за своїм функціональним призначенням дозволяють визначати положення людини на двовимірному зображенні на основі певних характерних ознак. Але ситуація ускладнюється, коли необхідно "оживити" зображення людини зі створенням відповідної точної 3D-фігури. При цьому все частіше впроваджують процедури навчання системи машинного зору з використанням комп'ютерних процедур риггінгу та текстурування. В роботі розглянуто базові математичні співвідношення, які дозволяють виділити особливості створення 3D-моделі людини з анімацією її руху на основі лише наявного статичного фотозображення. Показано, що ключова ідея при цьому полягає у тому, що потрібно використати шаблонний прототип 3D-моделі (SMPL-прототип) і далі просто необхідно контролювати, щоб ця обрана структура відповідала силуету на зображенні. Створення ж руху цієї моделі пропонується на основі генерування карт глибини, нормалі та скінінгу. При цьому слід відмітити, що точність руху моделі напряму залежить від створення ще і карти міток моделі тіла людини. Додатково в роботі наголошено, що на точність створення рухомої 3D-моделі на основі технології Photo wake-up, впливає точність пози моделі голови людини, напряму погляду, куту повороту голови, тощо. При цьому відмічено, що при використанні моделі SMPL через недостатню точність щодо орієнтирів обличчя на зображенні, рекомендовано при формуванні 3D-образу голови, оптимізувати в моделі розташування вершин, дотримуючись меж силуету, і уникаючи при цьому самоперетину елементів. Представлені в роботі підходи та відповідний математичний апарат дозволяють у майбутньому розробити удосконалену версію технології Photo wake-up для випадку, коли на площині 2D-зображення виділяється не один силует людини, а група людей, до якої також можна додати анімацію окремих рухів.

Ключові слова: *машинний зір, модель, скін, карта, деформація, меш, Photo Wake-Up.*

Постановка проблеми. В останні роки цифрова обробка і цифровий аналіз зображень знаходять все більше застосувань в різних галузях науки і техніки, серед яких варто виділити додатки систем доповненої реальності, інтелектуальні робото-технічні комплекси, системи промислового контролю, системи управління рухомими апаратами, біомедичні дослідження, нові технології обробки зображень і безліч інших. Термін «машинний зір» (Machine vision) як поняття, включає в себе ряд інженерних технологій, методів і алгоритмів, пов'язаних із завданнями інтерпретації сцени спостереження за її двовимірними проекціями (зображеннями), а також містить практичне використання результатів цієї інтерпретації. Проблематика машинного зору настільки приваблива для сучасних дослідників з тієї причини, що апаратні можливості, надані в даній області останніми досягненнями електроніки та обчислювальної техніки, досягли такого рівня, що вони вже багато в чому наближаються до «технічних характеристик» людини. Окремою ланкою досліджень можна вважати наявні розробки алгоритмів, які дозволяють анімувати об'єкти в нерухомих зображеннях, насамперед це стосується анімації об'єктів людей. Ключовою задачею цих алгоритмів є використання підходів машинного зору задля того, щоб електронна система з відповідним програмним забезпеченням дозволяла визначати положення людини на двовимірному зображенні з подальшим створенням її руху.

Постановка завдання. Визначення тривимірної структури за двовимірним зображенням значно ускладнюється, коли в якості об'єкту на цьому зображенні виступає людина. І тут задача не лише полягає у визначенні положення людини, його загальних розмірів, але й потрібно створити деталізовану реалістичну 3D-модель з анімацією. При цьому, необхідно щоб алгоритм працював в форматі доповненої реальності, а сама 3D-модель людини

мала б можливість анімації у трьох вимірах, створюючи ефект "виходу" з зображення. Подібними дослідженнями сьогодні займається Інститут інтелектуальних систем імені Макса Планка, де задля створення відчуття руху було запропоновано окремий спосіб оброблення 2D-зображення. За створенням у подальшому «інтелектуальних машин», що функціонують в реальному масштабі часу, стоїть необхідність вирішення здебільшого лише одного принципового завдання – розробки методів і алгоритмів «розуміння» зображень. Однак саме це завдання в багатьох випадках виявляється і найбільш важким.

Метою статті є розроблення алгоритму створення фото реалістичної 3D-моделі тіла людини з можливістю анімації її руху на основі лише оброблення 2D-зображення з образом цієї людини. Для досягнення поставленої мети необхідно виконати наступні завдання:

– виявлення базових математичних підходів, на основі яких функціонує система машинного зору з використанням технології Photo wake-up;

– визначення особливостей щодо використання SMPL-моделі, карт глибини, нормалей та скінінгу при створенні карти міток 3D-моделі;

– формулювання етапів створення 3D-моделі тіла людини з анімацією її руху на основі статичного 2D-зображення.

Виклад основного матеріалу дослідження. Однією з найбільш цікавих і популярних на сьогодні систем для створення руху об'єктів через аналіз плоского зображення, є анімація на основі установки, яка називається деформацією скелетного підпростору [1, с. 3]. Її часто використовують для управління особливо персонажами, але ця техніка універсальна і може бути використана до будь-якої 3D-моделі на основі мешу. Метод деформує вхідний меш шляхом застосування перетворень до простішої системи ригу, яку іноді називають скелетом. Риг складається з елементів, які називаються кістками, іноді їх також називають суглобами або ручками.

Величина впливу кожної кістки в контрольному ригу на дану вершину контролюється вагою кістки-вершини. Кінцева позиція v_j' кожної вершини j, з формули (1) задається зваженою сумою всіх кісткових перетворень

$$v_j' = \sum_i w_j^i T_i(v_j) \qquad (1)$$

де w_j^i – вага кістки i для вершини j і T_i є трансформацією кістки i [1, с. 3].

Риг, як проміжний допоміжний засіб, використовується в технології Photo wake-up, за якою, на основі 2D-зображення можна створити реалістичну динамічну 3D-модель. Загальна система на основі технології Photo wake-up працює наступним чином (рис. 1): спочатку застосовуються алгоритми для визначення персонажа, проводиться сегментація зображення та визначаються оцінки розташування об'єктів на зображенні. На основі результатів розробляється метод побудови ригід меш. Потім, будь-яку шаблонну 3D-послідовність руху (переміщення, обертання, зміна розмірів) можна використовувати для анімації ригід меш.

Виявлення персонажа, оцінка об'єкта та сегментація персонажа виконуються на основі фотографії, за допомогою алгоритмів нейронних мереж. Наприклад, візьмемо нейронну мережу Mask R-CNN для, по-перше, виявлення персонажу на зображенні, і по-друге, для його подальшого відокремлення від фону зображення. При цьому, для створення 3D-моделі використовують шаблонний прототип моделі людського тіла SMPL (Skinned Multi-Person Linear Model). Цей прототип визначає 3D-модель тіла, яка змінюється по формі фігури з вихідного двовимірного зображення. Додатково, на отриману модель, виконується процедура накладання текстур, тобто

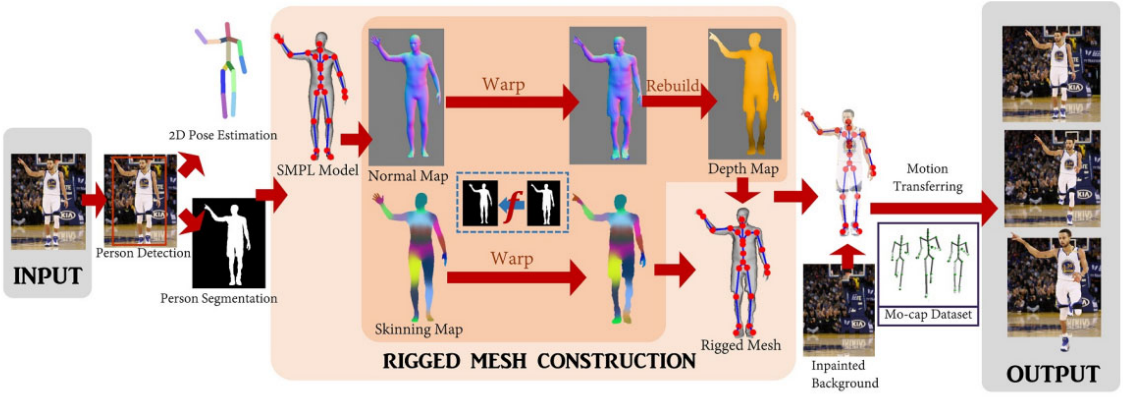

Рис. 1. Послідовність методу [2, с. 5910]

вони проектуються на зображення у вигляді карти нормалей і карти скінінгу. Суть системи – знайти відповідність між силуетом персонажа та силуетом SMPL, провести деформацію карт нормалей/скінів SMPL та створити карту глибини через інтеграцію викривленої карти нормалей. Цей процес циклічно повторюється в нейронній мережі, щоб імітувати задній вигляд моделі та поєднувати карти глибини та скінінгу для створення повного ригід 3D-моделі. Ключова ідея алгоритму у тому, як відновити анімовану текстуровану 3D-модель з однієї фотографії. Один із способів полягає у тому, щоб змусити модель SMPL відповідати силуетам, оптимізувавши в неї розташування вершин, дотримуючись меж силуету, і уникаючи при цьому самоперетину [3, с. 8389]. Отримані карти звичайних зображень і скінів можна побудувати як для переднього, так і заднього вигляду, а потім відновити ригід меш тіла.

Алгоритм починається з формування двовимірного образу персонажа та її силуетної маски *S*. Для простоти *S* називають і як набір усіх пікселів у силуеті, і як двійкову функцію $S(x) = 1$ для пікселя x всередині силуету, або $S(x) = 0$ для x поза силуетом.

Щоб побудувати 3D-модель зі скелетним оснащенням, спочатку підганяють модель SMPL до 2D-вхідного образу за допомогою методу, який описано в [4, с. 5–6], і який додатково відновлює параметри камери. Потім проектують цю модель в камеру, щоб сформувати силуетну маску S_{SMPL}. Проекція додатково дає карту глибини $Z_{\text{SMPL}}(x)$, карту нормалей $N_{\text{SMPL}}(x)$ і карту скінінгу $W_{\text{SMPL}}(x)$ для пікселів $x \in S_{\text{SMPL}}$. Відмітимо, що карта скінінгу отримана з ваг скінінгу для кожної вершини в моделі SMPL.

Керуючись S_{SMPL} і силуетною маскою *S* вхідної фотографії, деформують Z_{SMPL}, N_{SMPL} і W_{SMPL}, щоб побудувати вихідну карту глибини (лише на силуеті) $Z_{\partial S}(x \in \partial S)$, нормальну карту $N(x)$ і карта скінінгу $W(x)$, відповідно, для пікселів $x \in S$. $N(x)$ потім інтегрується для відновлення остаточної карти глибини $Z(x)$ за умови відповідності $Z_{\partial S}(x)$ на межі силуету ∂S. Тобто, розв'язують плавну зворотну деформацію, $f(x)$, так, як описано у співвідношенні (2):

$$S(x) = S_{\text{SMPL}}(f(x)) \tag{2}$$

а потім застосовують цю деформацію до карт глибини та спінінгу, відповідно до формул (3–6):

$$\text{Z}_{\partial S}(x \in \partial S) = \text{Z}_{\text{SMPL}}(f(x)) \tag{3}$$

$$N(x) = N_{\text{SMPL}}(f(x)) \tag{4}$$

$$Z(x) = \text{Integrate}[N; Z_{\partial S}] \tag{5}$$

$$W(x) = W_{\text{SMPL}}(f(x)). \tag{6}$$

Процедура викривлення зазвичай розтягує геометрію в площині (модель SMPL зазвичай є більш тонкою, ніж «одягнений» об'єкт), без аналогічного розтягування (як правило, збільшення) глибини. Вирішують цю проблему, натомість через деформацію нормалей, щоб отримати $N(x)$, а потім виконавши інтегрування їх, щоб отримати $Z(x)$ [5, с. 247].

Описаний метод відновлює карти глибини та скінінгу для передньої частини персонажа. Щоб відновити задню частину персонажу, віртуально візуалізують вигляд ззаду підігнаної моделі SMPL, віддзеркалюють маску персонажа, а потім застосовують метод деформації.

Реконструюють передній та задній меш стандартним способом: в глибину заднього проекту в 3D будують два трикутники для кожного сусіда 2×2. Кожній вершині призначають відповідні ваги скінінгу. Зшити передній та задній меш разом просто, оскільки вони збігаються на межі. На рисунку 2 показано передній та задній меш та зшиту модель.

Коли суб'єкт самозакривається, тобто одна частина тіла знаходиться поверх іншої, то реконструкцію однієї карти глибини (наприклад, для переду) із бінарного силуету буде недостатньо. Щоб впоратися з самооклюзією, сегментують тіло на частини за допомогою карти міток тіла, а потім реконструюють кожну частину за допомогою методу зіставлення меж. Зосереджуються на самооклюзії і коли руки моделі частково перетинають інші частини тіла так, що закриті частини залишаються єдиним з'єднаним компонентом. Цей метод не обробляє всі сценарії самооклюзії, але значно розширює робочий діапазон і показує шлях до обробки більшої кількості випадків.

Натомість, спроектована модель SMPL надає еталонну карту L_{SMPL} міток тіла, яка не дуже відпо-

(a) front mesh (b) side view (c) back mesh

Рис. 2. Реконструйовані результати мешу: a) реконструйований передній меш; b) вигляд збоку; c) реконструйований задній меш [2, с. 5912]

відає зображенню. Цю карту міток використовують для побудови остаточної карти міток L у два етапи: 1) оцінюють початкову карту міток L_{init} для кожного пікселя $x \in S$, щоб вона була максимально подібною до L_{SMPL}; 2) уточнюють L_{init} на границях оклюзії, де розриви міток повинні збігатися з краями вхідного зображення.

Початкова (приблизна) карта міток тіла L_{init} розраховується шляхом мінімізації мети випадкового поля Маркова (MRF), як це визначають формули (7–9).

$$\min_{L_{\text{init}}} \sum_{p \in S} U(L_{\text{init}}(p)) + \gamma \sum_{p \in S, q \in N(p) \cap S} V\left(L_{\text{init}}(p), L_{\text{init}}(q)\right) \quad (7)$$

де

$$U(L_{\text{init}}(p)) = \min_{r | L_{\text{SMPL}}(r) = L(p)} \|p - r\|_2 \quad (8)$$

$$V(L_{\text{init}}(p), L_{\text{init}}(q)) = \begin{cases} 1, & \text{якщо } L_{\text{init}}(p) \neq L_{\text{init}}(q) \\ 0, & \text{інакше} \end{cases} \quad (9)$$

Рисунок 3 ілюструє етапи відновлення карти міток. Починаючи з вхідного зображення (рисунок 3a) і його відповідного силуету та спроектованої моделі частини тіла SMPL, відновлюють початкову карту міток частини тіла, як показано на рисунку 3b. Після визначення точок на кордонах оклюзії будують маску оклюзії (світліші ділянки показано на рисунку 3c), а потім уточнюють її, щоб створити остаточну карту міток тіла (рисунок 3d). Частини тіла поблизу оклюзій мають фальшиві межі, і вони показані червоним кольором на рисунку 3e. Видаляють ці помилкові межі (рисунок 3f) і замінюють їх трансформованими версіями меж SMPL, як показано на рисунку 3g. Потім перебудовують тіло – частина за частиною (рисунок 3h) і збирають в остаточний меш, як це показано на рисунку 3i.

Точність пози голови за технологією Photo wake-up важлива для точної анімації, тоді як образ голови за SMPL часто є неправильним. Через це, на основі розробок [6 с. 1756, 7 с. 1750], виявляють орієнтири обличчя на зображенні та формують 3D-образ голови, який найкраще вирівнює відповідні, спроектовані 3D орієнтири з виявленими. Після реконструкції карти глибини для голови, як і раніше, застосовують плавну деформацію, яка точно вирівнює спроектовані 3D реперні точки з опорними точками зображення. Щоразу, коли обличчя чи реперні знаки не виявлено, цей крок пропускається.

Етап текстурування означає, що для передньої частини об'єкта проектують зображення на геометрію. Для задньої текстури пропонують два варіанти: 1) вставити дзеркальну копію передньої текстури на зворотний бік; 2) зафарбувати з додатковими вказівками користувача. Для другого варіанту розмальовування спини здійснюється за картами міток тіла, малюючи текстуру з регіонів з однаковими мітками тіла. Користувач може легко змінити ці мітки карти, щоб, наприклад, заохочувати заповнення потилиці текстурою волосся, а не текстурою обличчя. Нарешті, передня і задня текстури зшиваються пуассонівським змішуванням [8, с. 314]. На рисунку 4 наведено приклади карт міток тіла та сіток. На рисунку 4 оригінальні фотографії розміщено у верхньому правому куту набору результатів роботи технології Photo Wake-up.

Таким чином, алгоритм, за яким можна створити реалістичну 3D-модель тіла з анімацією для систем машинного зору можна визначити за наступними етапами:

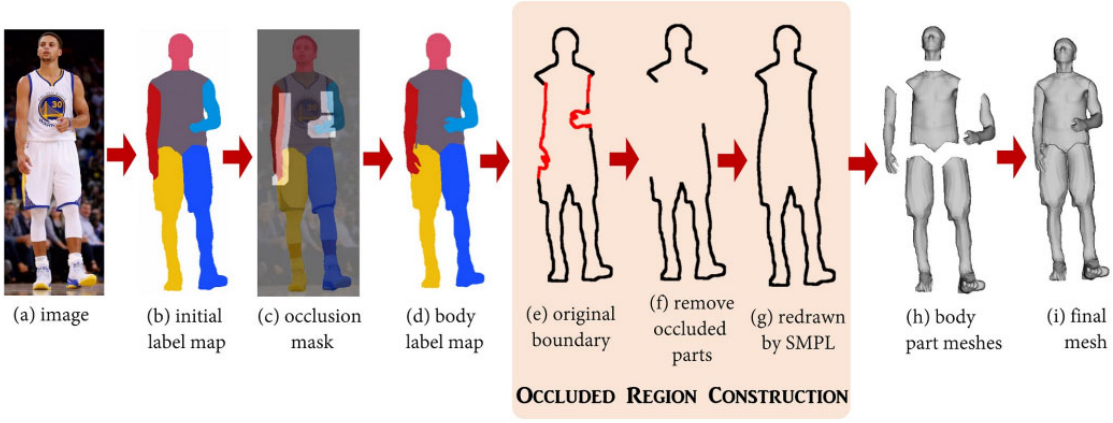

Рис. 3. Алгоритм відновлення карти міток: a) ілюстрація вхідного зображення; b) ілюстрація наближеної карти міток; c) ілюстрації маски оклюзії; d) ілюстрація остаточної карти міток; e) ілюстрація фальшивих меж оклюзій; f) ілюстрація видалених частин контуру; g) ілюстрація трансформованих меж SMPL; h) часткові складові моделі; i) ілюстрація остаточної моделі [2, с. 5913]

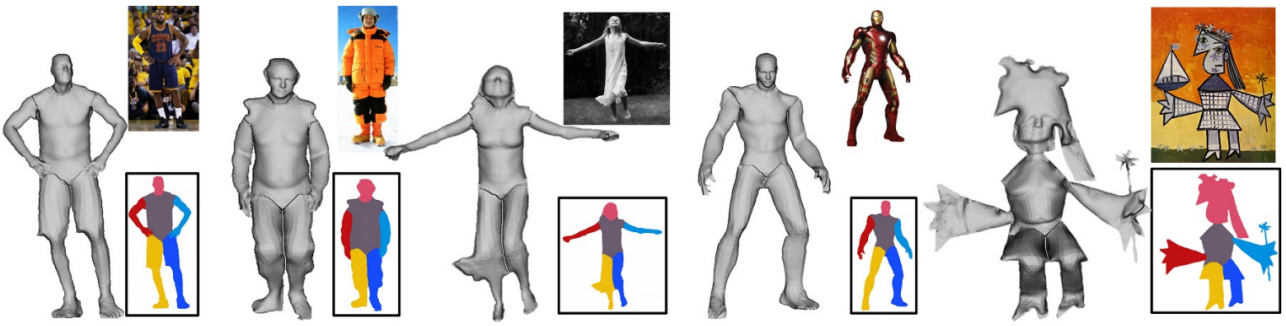

Рис. 4. Приклади карт міток тіла та сіток [2, с. 5914]

1) 2D-зображення ділиться на області за прикладом створення сітки;

2) Визначаються на основі прототипу SMPL, окремі частини тіла і змінюються так, щоб відповідати вихідному оригіналі на зображенні;

3) Проводиться корекція пози голови людини з врахуванням таких особливостей, як напрям погляду, кут повороту;

4) Проводиться на основі побудови ригу анімація тіла в трьох вимірах, змушуючи його "виходити" з площини плоского зображення.

Висновки. В роботі розглянуто базові математичні співвідношення, які дозволяють виділити особливості створення 3D-моделі людини з анімацією руху, на основі лише наявного статичного фотозображення. Показано, що ключова ідея при цьому полягає у тому, що потрібно використати шаблонний прототип 3D-моделі (SMPL-прототип) і далі просто необхідно контролювати, щоб ця обрана структура відповідала силуету на зображенні. Створення ж руху цієї моделі пропонується на основі генерування карт глибини, нормалі та скінінгу. При цьому слід відмітити, що точність руху моделі напряму залежить від створення ще і карти міток моделі тіла людини. Додатково, в роботі наголошено, що на точність створення рухомої 3D-моделі на основі технології Photo wake-up, впливає точність пози голови людини, напряму погляду, куту повороту голови, тощо. При цьому відмічено, що при використанні моделі SMPL через недостатню точність щодо орієнтирів обличчя на зображенні, рекомендовано при формуванні 3D-образу голови, оптимізувати в моделі розташування вершин, дотримуючись меж силуету, і уникаючи при цьому самоперетину елементів. Виділені в роботі підходи та відповідний математичний апарат дозволяють у майбутньому розробити удосконалену версію технології Photo wake-up для випадку, коли на площині 2D-зображення виділяється не один силует людини, а група людей, до якої також можна додати анімацію окремих рухів.

Список літератури:

1. J.P. Lewis, M. Cordner, and N. Fong, "Pose space deformation: a unified approach to shape interpolation and skeleton-driven deformation," Proceedings of the 27th annual conference on Computer graphics and interactive techniques, 165–172, 2000, doi: 10.1145/344779.344862

2. C. Weng, B. Curless, "Photo Wake-Up: 3D Character Animation From a Single Photo," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 5908-5917, 2020, doi: 10.1109/CVPR.2019.00606

3. T. Alldieck, M. Magnor, W. Xu, C. Theobalt, and G. PonsMoll, "Video based reconstruction of 3d people models," IEEE Conference on Computer Vision and Pattern Recognition, 8387-8397, 2018, doi: 10.48550/arXiv.1803.04758

4. F. Bogo, A. Kanazawa, C. Lassner, "Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image," Computer Vision – ECCV 2016, Lecture Notes in Computer Science, 1–18, 2016, doi: 10.48550/arXiv.1607.08128

5. R. Basri, D. Jacobs, and I. Kemelmacher, "Photometric stereo with general, unknown lighting," International Journal of Computer Vision, 72(3), 239–257, 2007, doi: 10.1007/s11263-006-8815-7

6. D. E. King. Dlib-ml, "A machine learning toolkit," Journal of Machine Learning Research, 10(7), 1755–1758, 2009.

7. I. Kemelmacher-Shlizerman, S. M. Seitz, "Face reconstruction in the wild. In Computer Vision (ICCV)," 2011 IEEE International Conference, 1746-1753, 2011, doi: 10.1109/ICCV.2011.6126439

8. P. Perez, M. Gangnet, and A. Blake, "Poisson image editing," ACM Transactions on graphics (TOG), 22, 313–318, 2003.

Trapezon K.O., Voinalovych O.O. FEATURES OF USING PHOTO WAKE-UP TECHNOLOGY FOR MACHINE VISION SYSTEMS

Machine learning systems, by their functional purpose, allow determining the position of a person on a two-dimensional image based on certain characteristic features. But the situation becomes more complicated when it is necessary to "animate" a person based on image analysis by creating a corresponding accurate 3D figure. At the same time, procedures for training the machine vision system using computer rigging and texturing procedures are increasingly being implemented. The paper considers the basic mathematical relationships that allow us to highlight the features of creating a 3D model of a person with motion animation based only on an existing static photo image. It is shown that the key idea here is that you need to use a template prototype of the 3D model (SMPL-prototype) and then you just need to control that this selected structure corresponds to the silhouette in the image. Creation of the movement of this model is proposed on the basis of the generation of depth, normal and skinning maps. At the same time, it should be noted that the accuracy of the movement of the model directly depends on the creation of the label map of the human body model. In addition, the work emphasizes that the accuracy of creating a moving 3D model based on Photo wake-up technology is affected by the accuracy of a person's head posture, direction of gaze, angle of head rotation, etc. At the same time, it was noted that when using the SMPL model, due to insufficient accuracy with respect to facial landmarks in the image, it is recommended to optimize the location of the vertices in the model when forming a 3D image of the head, observing the boundaries of the silhouette, while avoiding self-intersection of elements. The approaches identified in the work and the corresponding mathematical apparatus allow in the future to develop an improved version of the Photo wake-up technology for the case when not one silhouette of a person stands out on the plane of a 2D image, but a group of people, to which animation of individual movements can also be added.

Key words: *machine vision, model, skin, map, warp, mesh, Photo Wake-Up.*