

A. Shcherbakov, B. Moroz

MATHEMATICAL MODEL FOR OBTAINING A STEREOSCOPIC IMAGE FROM SEVERAL WIDE-ANGLE CAMERAS OF AN AIRCRAFT

Анотація. Актуальність теми дослідження не викликає сумніву, оскільки запропоновано вдосконалити систему стереоскопічного відеозображення. Тим самим буде синтезовано нову математичну модель отримання стереопанорами, яка приймає відеопотік із широкоформатних камер типу "риб'яче око". Така модель повина знизити витрати на панорамну відео систему, залишаючи високу якість вихідного зображення, без видимих швів склеювання фрагментів у відеоряді.

Мета роботи створення математичної моделі для отримання стереоскопічного зображення з декількох широкоформатних камер, зберігши при цьому вихідну якість відеоряду. Також, у цій роботі будуть запропоновані методи склеювання зображення та усунення дісторсії при синтезуванні панорамного зображення.

У роботі пропонується використовувати метод сферичної розгортки. У якому на зміну паралельним площинам використовуються концентричні віртуальні сфери з центром у системі координат системи. Пропонується система із двома ширококутними вертикально встановленими камерами типу "риб'яче око". Це дозволить отримати панорамну оцінку глибини шляхом проектування отриманих зображень із широкоформатних камер на чотири віртуальні площини, розташовані паралельно до базової лінії. Незважаючи на це, карти диспаратності властиво низька роздільна здатність через обмеження матриці та високої дісторсії об'єктива "риб'яче око", тому оцінка глибини вертикальних структур паралельних базової лінії, часто відсутня. Поряд з цим представлений алгоритм всеспрямованого рухаючогося стереозображення, який обчислює глибину статичної сцени в 360 градусів за коротким відеофрагментом, знятим обертаючоюся камерою.

Запропонована в цій роботі математична модель має більш ефективні та менш економічно затратні параметри для створення стереопанорами із оглядом 360 градусів. Для створення стереопанорами була представлена ефективна система реєстрації відеозображення, в якій використовувалося всього чотири широкоформатні камери з перетинаючимися кутами огляду, для захоплення відео. Точне розміщення камер дозволило відкалібрувати систему для склеювання панорамного зображення.

Ключові слова: FOV, стереопанорама, об'єктив «риб'яче око», спотворення, стереозображення, 360° відео, VR, калібрування.

Introduction and formulation of the research problem. Recently, video with a 360° panoramic view has attracted more and more attention from developers, as it opens up completely new prospects for the implementation of VR technologies. Such giant companies as Samsung, Sony, HTC are actively improving their technologies, which can be seen in the growth of product trends every year. Also, VR equipment has become more affordable for the end-user, and simple VR system setup has become a key moment in the growth of interest among the community. Not so long ago, several systems that allow you to create 360-degree panoramas have been demonstrated. These systems showed satisfactory quality of the stereo panoramic image. However, it should be noted that the proposed systems are based on expensive equipment. In the proposed architecture, the cameras are installed in a circle at a short distance from each other. Thereby, they add unnecessary load during data transmission and processing, and also greatly increase the cost of the structure. In this article, it is proposed to improve the stereoscopic video image system. Thus, a new mathematical model for obtaining a stereo panorama will be synthesized. It will receive a video stream from large-format “fisheye” cameras. Such a model is designed to reduce the cost of a panoramic video system, while leaving a high quality of the output image, without visible seams of combining fragments in the video sequence.

Goal. To propose a mathematical model for obtaining a stereoscopic image from several large-format cameras, while maintaining the original quality of the video sequence. Also, in this work, methods of image combining and distortion elimination when synthesizing a panoramic image will be proposed.

Analysis of recent research and publications. At the moment there are several scientific developments concerning our topic:

1. C. Forster et al. proposed an algorithm for visual monocular odometry based directly on pixel intensity, providing sub-pixel accuracy at high frame rates [1].
2. D. Scaramuzza et al. developed a procedure for calibrating a wide-format camera from a fixed point of view [2].
3. D. Scharstein et al. considered the usage of panoramic images in SFM algorithms. A new method of using spherical image models on catadioptric and dioptric systems was proposed [3].
4. W. Miled et al. described a new variational method for estimating the disparity of stereo images based on a block iterative model, that allows you to easily include a wide range of constraints, taking advantage of parallel computing architectures [4].

Having considered the proposed mathematical models, a general idea of the operation of a mono-stereo system was obtained. Based on the data obtained, it is proposed to synthesize its own optimally balanced stereo vision system, which allows monitoring the surrounding space in real time using VR technologies.

Presentation of the main material. There are several basic approaches to creating panoramic images. The first and easiest way to create a panoramic image is to use a single camera. The method is to project the image display through a mirror onto the camera's matrix. This method is quite simple to use, but the resulting image is of extremely low quality. Another technique makes up for the limitations of hardware by generating a panoramic image in software. This technique evaluates camera movement by continuously monitoring the scene being shot and combines the captured images together. Despite the simple approach and high quality of the output image, this method is susceptible to geometric collisions, color distortion and adding artifacts when improperly captured. The last method, one of the most expensive to implement, suggests using several cameras located in a circle. This method involves fine-tuning the position of the cameras relative to each other so that two adjacent cameras overlap the capture area. After receiving video from individual cameras, the fragments are combined together using specially developed algorithms.

Quite a large contribution was made to the research and development of mono panoramic systems. Mono panoramic scenes are a type of image that is shot from different viewing angles and then combined together to create a wide-angle panoramic image. At the moment, most of the 360-degree panoramas that are openly available on the Internet are mono panoramas. Mono panorama looks the same for the left and right eyes and does not provide information about the depth and volume of the scene being shot. Such a system is difficult to shoot and takes a long time to create 360-degree panoramic images. Mono panoramas have been replaced by stereo panoramas. Stereo system - consists of two separate images for the left and right eyes with a slight horizontal offset. This technique involves shooting with two independent cameras. This approach, based on taking images from two different viewing angles, creates the illusion of depth in the image. The same principle is embedded in human vision, where the brain, receiving binocular images, can give a personal assessment of the distance and size of the objects under consideration, and is also able to synthesize a large-format image from two fields of view overlapping in the center of the eyes. To create a stereo panorama expensive equipment, high processing power and long rendering time of footage are required, because it is necessary to create two separate panoramas for both eyes.

In stereo systems, the surface sweep method which is traversed over parallel virtual planes and projected onto the plane to search for stereo matches is used. Gallup [5] proposes an innovative, more robust method that suggests the use of multiple directional sweeps. Over time, this method has been expanded for use with “fisheye” lenses. In this work, it is proposed to use the spherical sweep method, similar to the Im method [6]. In this method, instead of parallel planes, concentric virtual spheres with a center in the coordinate system of the system are used. In omnidirectional stereo systems, it is common to use mirrors with multiple cameras. The Geiger [7] and Schonbein [8] algorithms use two cameras with mirror-lens optical lenses with a viewing angle of 360 ° are used. Despite the fact that they are capable of creating a panoramic stereo image, they have a significant disadvantage in the form of two blind zones along the epipolar plane, along which it is difficult to assess the depth due to its inconsistency. Gao and Shen [9] propose a system with two wide-angle, vertically mounted “fisheye” cameras. This allows us to get a panoramic depth estimate by projecting acquired images from wide-angle cameras onto four virtual planes parallel to the baseline. Despite this, disparity maps are characterized by low resolution due to the limitations of the matrix resolution and high distortion of the “fisheye” lens, therefore, estimates of the depth of vertical structures parallel to the baseline are often absent. Along with this, the Im [6] omnidirectional moving stereo image algorithm is presented, which calculates the depth of a static scene of 360 degrees from a short video clip captured by a rotating camera. The sphere deployment method allows obtaining images from any angles, which removes the restriction on stationary equipment placement.

According to Scharstein [10], there are four steps of estimating the depth of a stereopair: calculating the initial cost of alignment, aggregating the cost, calculating and optimizing the disparity map, and improving the disparity map. The most complex and difficult part of these steps is calculating the alignment cost from the input images. Typical costs for aligning a stereopair are the sum of absolute differences (SAD), normalized cross-correlation (NCC), rank transform, etc. Hirschmuller [11] compares and evaluates these alignment cost functions. Instead of looking for the minimum values of locally aggregated alignment costs, the depth map can be computed by global optimization using graph splitting, but they require a lot of computing power. Semiglobal matching (SGM) proposed by Heiko Hirschmüller [11] is an efficient way to globally share costs using dynamic programming.

We use the omnidirectional camera model, $x = \Pi(X; \Phi)$, where Φ is the parameters of the “fisheye” lens. The normalized x coordinate of the image is convert-

ed to pixel coordinates using a special transformation $A(x)$. To calibrate installed cameras, the traditional checkerboard method is used, where the internal and external parameters of the lens are calculated for each camera, the positions relative to the checkerboard are determined and finally all external and internal parameters are optimized. A large checkerboard is used to provide sufficient overlap between the camera views. The outer parameters are represented as $= (r^T, t^T)^T$, where r is rotation vector of the axis-angle, t is translation vector ($r, t \in R^3$). Rigid Transformation Matrix $M(\theta)$ is given as $[R(r) \ t]$, where $R(r)$ is the 3×3 , rotation matrix corresponding to r . Based on the images of the chessboard of camera i , we denote its lenses by Φ_i and $A_i, \{\theta_i, k\}$, where k is the capture index. The relative exposure from camera i to j can be calculated as $\theta_{j,k} * \theta_{i,k}^{-1}$ from a pair of simultaneously obtained images (i, k) and (j, k) , where $*$ and $^{-1}$ mean composition and inversion operations.

or external calibration, all positions of the camera $\{\theta_i\}$ and the positions of chess cells $\{\theta_k\}$ are initialized in the coordinate system of the first camera, thereby we minimize the error when projecting points in the corners of the chessboard (1).

$$\sum_{(i,k)} \sum_p \|\tilde{x}_{i,p} - A_i \left(\Pi(M(\theta_i * \theta_k) [X_p \ 1]; \Phi_i) \right)\|, \quad (1)$$

where $\{(i, k)\}$ is the data set of measurements of the chessboard from position k using camera i , X_p is the coordinate of the corner point p on the chessboard, and $\tilde{x}_{i,p}$ is the pixel coordinate of the X_p in image i . The Ceres algorithm [12] is used for optimization.

The sweep line algorithm, [13] provides tight stereo alignment between multiple images. However, this algorithm is difficult to apply to images taken from wide-angle cameras with a view of more than 180 degrees. Svoboda [14] uses several planes with different normals and distances, and later Im [6] uses local spheres centered relative to one camera to estimate the depth of the camera's spherical panorama. The proposed system works with wide-format stereo images, which existing algorithms cannot do. To estimate depth in a panoramic stereo system, an algorithm for estimating spherical depth is proposed.

The center of the sweep can be in an arbitrary place, but to reduce the distortion of spherical images, select the center of the system as the origin and align the xz plane so that it is located close to the centers of the cameras. In this spherical coordinate system, the ray $p = (\theta, \varphi)$ corresponds to $\rho(p) = (\cos(\varphi)\cos(\theta), \sin(\varphi), \cos(\varphi)\sin(\theta))^T$. Let the transformed external parameters of the camera in the setup coordinate system be $\{\theta_i^*\}$. Also, for simplicity, we

denote the projection function $A_i(\Pi(X; \Phi_i))$ by $\Pi_i(X)$. We are now projecting the input images onto virtual spheres. Each pixel in the resulting spherical image S represents a ray (θ, φ) . The spherical image S has a resolution of $W \times W$, and θ ranges from $-\pi$ to π . φ can take values from $-\pi/2$ to $\pi/2$. N (number) of spheres are chosen so that their reciprocal depths are uniform, i.e. when the minimum depth is D_{min} , the inverse depth to the n th sphere is $d_n = n/(D_{min}(N - 1)), n \in [0, \dots, N - 1]$. That is, the radii of the spheres are equal to $1/d_n$, with the exception of $n = 0$, which corresponds to an infinite sphere.

The pixel value of the spherical image is calculated by an equation of the form (2).

$$S_{i,n}(p) = I_i \left(\Pi_i \left(M(\theta_i^*) \left[\rho(p)/d_n \right] \right) \right) \quad (2)$$

where I_i is the input image received by camera i . For $n = 0$ we use $d_0 = 2^{-23}$. If the projected pixels do not fall into the visible area of the input image, we do not take them into account in further processing.

In spherical sweep algorithms, we need to compute the $W \times H \times N$ volume cost C for all rays and reciprocal depths. Suppose we are given a paired compatibility cost function $F(*, *)$, that takes two images and computes a cost map of the same size. The integrated cost map is the average of all possible (and valid) paired cost maps, and the cost volume is the set of integrated cost maps, i.e. the cost p on the n -th sphere (3).

$$C(p, n) = \{ F(S_{i,n}(p), S_{j,n}(p)) \} \quad (3)$$

where i, j is an unordered index pair of spherical images.

The resulting volume often contains noises, for its refinement we use the advantages of SGM [15], by minimizing the energy function using dynamic programming. Finally, the inverse ray depth p is defined as d_{n^*} , where $n^* = \operatorname{argmin}_n C(p, n)$.

Conclusion and prospects for further development. Unlike other existing systems for creating panoramic video images, the mathematical model proposed in this work has more efficient and less cost-effective parameters for creating a stereo panorama with a 360 degree view. The proposed method for creating a stereo panorama was fully automated and does not require external intervention. To create a stereo panorama, an efficient video capture system was introduced, using only four wide angle cross-view cameras to capture video. The precise placement of the cameras allowed the system to be calibrated for combining the panoramic image. In ad-

dition, the resulting panorama eliminates the presence of visible seams and is automatically straightened. The proposed model prevails over the previously created systems for creating stereo content both from a qualitative and quantitative point of view.

REFERENCES

1. C. Forster. Fast semi-direct monocular visual odometry [Text] / C. Forster, M. Pizzoli, and D. Scaramuzza // In IEEE Int. Conf. on Robotics and Automation. – 2014. – P. 15–22.
2. D. Scaramuzza. A flexible technique for accurate omnidirectional camera calibration and structure from motion [Text] / D. Scaramuzza, A. Martinelli, and R. Siegwart / Proceedings of the Fourth IEEE International Conference on Computer Vision Systems. – 2006, P. 45–45
3. D. Scharstein. Structure from motion using full spherical panoramic cameras [text] / D. Scharstein. A. Pagani and D. Stricker // In Int. Conf. on Computer Vision Workshops. –2011. P. 375–382.
4. W. Miled. Disparity map estimation using a total variation bound [Text] / W. Miled, J. C. Pesquet and M. Parent // Canadian Conf. Computer Robot Vision. –2006. P. 48–55.
5. D. Gallup. Real-Time Plane-Sweeping Stereo with Multiple Sweeping Directions [Text] / D. Gallup, J. Frahm, P. Mordohai, O. Yan, M. Pollefeys // Proceedings IEEE Conference on Computer Vision and Pattern Recognition. – 2007. – P. 1–8.
6. S. Im. All around depth from small motion with a spherical panoramic camera [Text] / S. Im, H. Ha, F. Rameau, H. Jeon, G. Choe, I. Kweon // Proceedings IEEE Conference on Computer Vision. – 2016. – P. 156–172.
7. C. Geyer. Catadioptric Projective Geometry [Text] / C. Geyer, K. Daniilidis. International Journal of Computer Vision. – 2001. – P. 223–243.
8. M. Schonbein. Omnidirectional 3d reconstruction in augmented manhattan worlds [Text] / M. Schonbein, A. Geiger. Proceedings IEEE Intelligent Robots and Systems. – 2014. – P. 716–723.
9. W. Gao. Dual-fisheye omnidirectional stereo [Text] / W. Gao, S. Shen. Proceedings IEEE Intelligent Robots and Systems. – 2017. – P. 6715–6722.
10. D. Scharstein. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms [Text] / D. Scharstein, R. Szeliski, R. Zabih // Proceedings IEEE Workshop on Stereo and Multi-Baseline Vision. – 2001. – P. 131–140.
11. H. Hirschmuller. Evaluation of Stereo Matching Costs on Images with Radiometric Differences [Text] / H. Hirschmuller, D. Scharstein // Proceedings IEEE

Transactions on Pattern Analysis and Machine Intelligence. – 2009. – P. 1582-1599.

12. C. Forster. Semidirect Visual Odometry for Monocular and Multicamera Systems [Text] / C. Forster, Z. Zhang, M. Zhang, M. Werlberger, D. Scaramuzza // Proceedings IEEE Transactions on Robotics. – 2016. – P. 249–265.

13. Sweep line algorithm. [Electronic resource] // URL: https://en.wikipedia.org/wiki/Sweep_line_algorithm.

14. T. Svoboda. Epipolar geometry for panoramic cameras [Text] / T. Svoboda, T. Pajdla, and V. Hlavac // In IEEE Conference on Computer Vision and Pattern Recognition. – 1998. P. 218–231.

15. H. Hirschmuller. Stereo processing by semiglobal matching and mutual information [Text] / H. Hirschmuller // Proceedings IEEE Transactions on pattern analysis and machine intelligence. – 2008. – P. 328–341.

Received 05.01.2022.

Accepted 10.01.2022.

Математична модель отримання стереоскопічного зображення з декількох ширококутних камер літального апарату

Аналіз останніх досліджень та публікацій. На даний момент існують кілька наукових розробок щодо нашої теми:

1. C. Forster та ін. запропонували алгоритм візуально монокулярної одометрії, який працює безпосередньо з інтенсивністю пікселів, що забезпечує субпіксельну точність за високої частоти кадрів.

2. D. Scaramuzza та ін. розробили методику калібрування широкоформатної камери з фіксованої точки огляду.

3. D. Scharstein та ін. розглянули використання панорамних зображень у SFM алгоритмах. Запропоновано новий метод використання сферичних моделей зображення на катадіоптричних та діоптричних системах.

4. W. Miled та ін описали новий варіаційний метод оцінки диспаратності стереозображень побудований на блочно ітеративній моделі, яка дозволяє легко включати широкий спектр обмежень, використовуючи переваги паралельних обчислювальних архітектур.

Розглянувши запропоновані математичні моделі, отримано загальне уявлення про боти моно стерео системи. На основі отриманих даних пропонується синтезувати свою оптимально збалансовану систему стерео бачення, що дозволяє моніторити навколишній простір у режимі реального часу із застосуванням VR технологій.

Мета. Запропонувати математичну модель отримання стереоскопічного зображення з кількох широкоформатних камер, зберігши при цьому вихідну якість відеоряду. Також, у цій роботі будуть запропоновані методи склеювання зображення та усунення дісторсії при синтезуванні панорамного зображення.

Викладення основного матеріалу. Існує декілька основних підходів до створення панорамних зображень. Перший і найлегший спосіб створення панорамного зображення має на увазі використання однієї камери. Метод полягає в тому, щоб проектувати зобра

ження через дзеркало на матриці камери. Цей спосіб досить простий у застосуванні, однак отримане зображення має вкрай низьку якість. Інша техніка перемагає обмеженість апаратних засобів, генеруючи панорамне зображення через програмне забезпечення. Така методика оцінює рух камери шляхом безперервного моніторингу сцени, що знімається, і зшиває зроблені зображення воедино. Незважаючи на простий підхід і високу якість вихідного зображення, цей підхід піддається геометричним колізіям, спотворенню кольору та появі артефактів при неналежній зйомці. Останній метод, один із найзатратніших у реалізації, пропонує використовувати кілька камер, розташованих по колу. Цей метод передбачає досить точне налаштування положення камер щодо один до одного таким чином, щоб дві сусідні камери перекривали захоплювану область. Після отримання відеоряду з окремих камер фрагменти зшиваються разом за допомогою спеціально розроблених алгоритмів.

У стереосистемах використовується метод розгортки поверхні, який пробігає по паралельним віртуальним площинам і проектується на площині для пошуку стереовідповідностей. Gallur пропонував інноваційно інший, надійніший метод, який спорядковує використовувати множино спрямовані розгортки. Згодом цей метод був розширений для застосування його з об'єктивами типу "риб'яче око". У цій роботі пропонується використовувати метод сферичної розгортки, аналогічний методу It. В якому на заміну паралельним площинам використовуються концентричні віртуальні сфери з центром в системі координат системи. У всепрямованих стереосистемах поширене використання дзеркал з кількома камерами. В алгоритмах Geiger та Schonbein використовуються дві камери із дзеркально лінзовими оптичними об'єктивами з кутом огляду 360°. Незважаючи на те, що вони здатні створювати панорамне стереозображення, вони мають суттєвий недолік у вигляді двох сліпих зон вздовж епіполярної площини, вздовж якої складно провести оцінку глибини через її непостійність. Gao та Shen пропонують використовувати систему з двома ширококутними вертикально встановленими камерами типу "риб'яче око". Це дозволяє отримати панорамну оцінку глибини шляхом проектування отриманих зображень з широкоформатних камер на чотири віртуальні площини, розташовані паралельно до базової лінії. Незважаючи на це, картам диспаратності властиво низька розподільна здатність через обмеження розподільної здатності матриці і високої дісторсії об'єктива "риб'яче око", тому оцінка глибини вертикальних структур паралельних базової лінії, часто відсутні. Поряд з цим представлений рухаючийся алгоритм всепрямованого стереозображення It, який обчислює глибину статичної сцени в 360 градусів за коротким відеофрагментом, знятим обертається камерою. Метод розгортання сфери дозволяє отримувати зображення у будь-яких ракурсах, що знімає обмеження на нерухоме розміщення обладнання.

Згідно Scharstein, існує чотири етапи оцінки глибини стереопари: обчислення по частковій вартості суміщення, агрегування вартості, обчислення та оптимізація карти диспаратності, покращення карти диспаратності. Найбільш складною і складною частиною цих етапів є обчислення вартості суміщення за вхідними зображеннями. Типові витрати на суміщення стереопари є сумою абсолютних різниць (SAD), нормалізовану перекресну кореляцію (NCC), рангове перетворення itd. Hirschmuller порівнює та оцінює ці

функції вартості суміщення. Замість пошуку мінімальних значень локально агрегованих цін суміщення, карта глибини може бути обчислена шляхом глобальної оптимізації за допомогою поділу графа, але вони вимагають великих обчислювальних потужностей. Напівглобальна відповідність (SGM), запропонована Hirschmuller це ефективний спосіб глобального поєднання витрат за допомогою динамічного програмування.

Ми використовуємо модель всеспрямованої камери, яка визначає дисторсію об'єкти в поліноміальною функцією.

Висновок. На відміну від інших вже існуючих систем для створення панорамного відеозображення, запропонована математична модель має більш ефективні та менш економічно затратні параметри для створення стерео панорами з оглядом 360 градусів. Запропонований метод для створення стереопанорами був повністю автоматизований і не вимагає стороннього втручання. Для створення стереопанорами була представлена ефективна система для реєстрації відеозображення, в якій використовувалося всього чотири широкоформатних камери з перетинаючимися кутами огляду, для захоплення відео. Точне розміщення камер дозволило відкалібрувати систему для склеювання панорамного зображення. Крім того, отримана панорама виключає наявність видимих швів та автоматично випрямляється. Запропонована модель переважає раніше розроблені системи для створення стерео контенту як з якісної, так і з кількісної точки зору.

Щербаков Артем Геннадійович - аспірант кафедри програмного забезпечення комп'ютерних систем Національного технічного університету «Дніпровська політехніка».

Мороз Борис Іванович - доктор технічних наук, професор кафедри програмного забезпечення комп'ютерних систем Національного технічного університету «Дніпровська політехніка».

Shcherbakov Artem - PhD student of the Department of Software Engineering of University of Technology.

Boris Moroz - Doctor of Technical Sciences, Professor, Head of the Department of Software Engineering of University of Technology.