Paper Title\* **(**use style: ***paper title***)

Subtitle as needed ***(paper subtitle)***

Authors Name/s per 1st Affiliation (*Author*)

line 1 (of *Affiliation*): dept. name of organization

line 2-name of organization, acronyms acceptable

line 3-City, Country

line 4-e-mail address if desired

Authors Name/s per 2nd Affiliation (*Author*)

line 1 (of *Affiliation*): dept. name of organization

line 2-name of organization, acronyms acceptable

line 3-City, Country

line 4-e-mail address if desired

*Abstract*— In response to the advancements in artificial intelligence (AI) technology, motion capture systems have become more accessible to small content creation industries due to reduced production costs and complexity. However, AI-based motion capture systems that rely on estimating motion without conventional markers may encounter accuracy issues. To address this challenge, we propose a low-cost and straightforward alternative, employing color-based markers and a multi-camera setup to capture actor movements. The primary objective of this study is to develop a reliable multi-camera system capable of estimating 3D coordinates using triangulation techniques. The system's performance is assessed through the reprojection error metric, which serves as an indicator of the accuracy of 3D estimation. Extensive testing involving four distinct camera rotation ranges reveals that reprojection error can be effectively employed as a measure of the reliability of 3D estimation results. Furthermore, our analysis demonstrates that camera rotation has minimal impact on reprojection error, while utilizing three synchronized cameras for triangulation yields the most optimal results with the lowest reprojection error and the best 3D estimation outcomes.

Keywords— Keywords: Motion capture, Multi-camera system, Triangulation, Reprojection error, 3D estimation.

# Introduction (*Heading 1*)

The progress in artificial intelligence (AI) technology has significantly enhanced the practicality and affordability of motion capture systems. However, the accuracy of motion capture results obtained from AI-based systems remains questionable. On the other hand, conventional motion capture systems offer precise and clear motion capture results but entail high costs due to the required materials, equipment, and skilled technicians, making them generally inaccessible for small industries. Despite their superior accuracy compared to AI-based systems, the high costs associated with conventional systems prohibit their replication and performance comparison in this study. Thus, to facilitate a fair comparison with AI-based motion capture systems, a non-AI-based motion capture system that can approximate the performance of conventional systems without relying on expensive equipment and materials is essential. This study focuses on discussing the methodology of estimating the 3D positions of markers on actors' bodies using a multi-camera system. However, estimating 3D positions from 2D images is not a straightforward task as valuable 3D information is lost during the image capture process.

# METHODOLOGY

## Image Formation in Cameras

To map a 3D point in the world onto a 2D point in a camera, transformations are employed, as described in Equation II.6. These transformations involve extrinsic matrices, which map 3D points from the world coordinate system to 3D points in the camera coordinate system, and intrinsic matrices, which map 3D points from the camera coordinate system to 2D points in the image coordinate system. Given the 2D points in the image coordinate system and the values of each camera's intrinsic and extrinsic matrices, it is possible to perform the inverse projection, allowing the 2D points in the image coordinate system to be reconstructed as 3D points in the world coordinate system. However, a challenge arises as information regarding the intrinsic and extrinsic matrices of the camera is often unknown or not readily available in the cameras used for this study.

## Camera Calibration (Single Camera)

Camera calibration, also referred to as single camera calibration in this study, involves the process of determining the intrinsic and extrinsic matrices. The calibration method utilized in this research is the Zhang Calibration Method, as discussed in subchapter II.3. This method allows for the estimation of the camera's intrinsic matrix and distortion parameters using multiple calibration pattern images, such as a known-sized chessboard with known numbers of rows and columns and the size of each square on the chessboard pattern. The distortion parameters represent image distortions caused by the camera lens and are crucial for distortion correction. Eliminating distortion is essential to enhance the accuracy of subsequent processes, as elaborated in subchapter II.4. Distortion correction techniques are further discussed in subchapter II.4.

However, the Zhang Calibration Method has a limitation; it does not provide the extrinsic matrix, which maps points from the world coordinate system to the camera coordinate system, hindering the back-projection of 2D image coordinates to 3D world coordinates. Nevertheless, since the intrinsic matrix has been obtained through the Zhang Calibration process, it is possible to perform the back-projection from the 2D image coordinate system to the 3D camera coordinate system, thereby recovering the lost 3D information during image capture.

## Triangulation for 3D Point Reconstruction

As discussed in the early part of subchapter II.5, the back-projection of 2D image coordinates to 3D camera coordinates from a single viewpoint may result in unclear 3D projections, particularly concerning the depth information or the Z-axis value of a 3D point. This ambiguity is due to an infinite number of 3D points satisfying the mapping equation from a 3D point to its 2D image point, as illustrated in Figure II.6. However, employing more than one view that observes the same 3D point yields at least one 3D point that satisfies the image mapping equation from each view, as demonstrated in Figure II.7.

The process of determining the 3D point using multiple views is known as triangulation. As explained in subchapter II.5, triangulation essentially utilizes fundamental linear algebra concepts, seeking the intersection points of two or more lines, where the intersection points in this context represent the sought-after 3D points. The projection function is the result of multiplying the intrinsic and extrinsic matrices, as given by Equation II.14. As at least two non-parallel lines are required to find the intersection point, triangulation necessitates a minimum of two views observing the same 3D point. By combining the projection matrices of each camera and the corresponding 2D points of the observed 3D point from each camera, a matrix A is constructed, as shown in Equation II.20. The solution of Equation II.22 will yield the 3D intersection point of the projection lines from each camera.

In real-world scenarios, obtaining the 3D intersection point from the individual camera back-projections is challenging since the back-projection lines of each camera do not typically intersect due to noise introduced from the observed 2D points or the camera projection matrices. Therefore, the problem of finding the 3D intersection points from the camera projection lines is transformed into a problem of searching for the best 3D point that can represent the intersection of the projection lines. To address this issue, the Singular Value Decomposition (SVD) method is commonly applied, performing a decomposition of matrix A to obtain the best-fitting 3D point that approximates the intersection of the camera projection lines.

## Stereo Calibration for Extrinsics Estimation

One unresolved issue in the previous stages of the process arises from the fact that the projection matrix utilized in triangulation is derived from the multiplication of the intrinsic and extrinsic matrices of each camera. The problem lies in assuming that the extrinsic matrix used in the triangulation process transforms the camera coordinate system to the world coordinate system, while this extrinsic matrix has not been obtained up to this point. An alternative solution is to perform triangulation operations in the context of one camera's coordinate system, rather than transforming the coordinate systems of each camera to the world coordinate system. This approach involves finding a transformation matrix that maps the coordinate system of one camera to the coordinate system of the target camera. This search for the transformation matrix is more feasible compared to obtaining the extrinsic matrices of each camera and is known as stereo calibration.

Similar to the extrinsic matrices that perform rotation and translation from the world coordinate system to the camera coordinate system, stereo calibration seeks a matrix that performs rotation and translation, transforming the coordinate system of one camera to the coordinate system of another camera. Utilizing the stereo calibration technique described in subchapter II.6, which involves using a calibration pattern such as a chessboard pattern visible to both cameras simultaneously, the required rotation and translation matrices can be obtained.

## Multi-View Triangulation for 3D Point Estimation

At this stage, the intrinsic matrices and the rotation-translation matrices from one camera to another have been acquired. The information obtained thus far is sufficient to proceed with 3D point estimation using triangulation techniques.

When considering the structure of the 3D point estimations obtained through triangulation, employing more than two camera views can significantly improve the quality of the resulting 3D point estimates. The use of multiple views enhances the robustness of the 3D point estimation process against noise present in the utilized parameters, leading to reduced errors in the estimation. Further details regarding 3D point estimation with more than two views are discussed in subchapter II.7.

## Multi-View Triangulation for 3D Point Estimation

The final stage, which is optional but considered a crucial aspect and a standard practice in 3D point estimation, involves optimizing the results obtained from the triangulation process. In this stage, a commonly used method called bundle adjustment is applied to optimize the projection matrices and the 3D points obtained from the triangulation process. The bundle adjustment aims to minimize the errors in the back-projection of 3D points to their corresponding 2D points from each camera used. Further details on the 3D optimization through bundle adjustment can be found in subchapter II.9.

After completing the optimization stage, it is anticipated that the optimized 3D point estimation results will be more accurate, as errors in the back-projection of 3D points to their 2D image points from each camera are minimized through the bundle adjustment process.

# Experiment

## Camera Calibration and Distortion Correction

In this experiment, the evaluation of both the camera calibration and distortion correction processes is presented in a unified subchapter. The camera calibration process begins with capturing calibration pattern images for each camera. The number of calibration pattern images, using a chessboard pattern, ranges from 10 to 20 images as per Zhang's guidelines (Zhang, 1998). After detecting the edges within the calibration patterns, the 2D detection points and their corresponding 3D positions, as listed in Table III.1, are fed into the cv.calibrateCamera function, resulting in the intrinsic camera matrices and distortion parameters.

One consistent finding during image capture is that the initial images, taken before applying distortion correction, do not exhibit noticeable distortion. This observation leads to the assumption that the inherent distortion in the cameras, if present, is not perceivable to the naked eye. This can be attributed to the use of lenses with relatively minor curvature, resulting in insignificant distortion. Consequently, a new assumption is drawn that the distortion correction process should not substantially alter the original images due to the small magnitude of initial distortion.

However, in the early stages of the experiment, this assumption was not met. The distortion correction process not only failed to eliminate minor distortion but also introduced significant distortion, as illustrated in Figure IV.1 below.

Figure IV.1. (a) Original image before distortion correction. (b) Image after distortion correction.

After conducting several iterations of the experiment, two factors contributing to this issue were identified:

* The calibration pattern images were not uniformly distributed across the camera's field of view, primarily concentrated in the center. This caused the calibration algorithm to model distortion parameters primarily in the central region, neglecting other areas.
* The calibration algorithm utilized a low degree of distortion coefficients. According to OpenCV documentation, the default degree of distortion coefficients is set to 3 out of a total of 6 available degrees. During the experiments, it was evident that the use of only three degrees was inadequate in accurately modeling the distortion.

Upon rectifying these errors by capturing calibration pattern images uniformly across the camera's view and increasing the distortion coefficient degree from 3 to 6, the distortion correction produced results similar to the original images, as shown in Figure IV.2 below.

Figure IV.2. (a) Original image before distortion correction. (b) Image after distortion correction.

In this experiment, no further testing was conducted on other available distortion coefficient degrees in OpenCV, as the improved distortion correction results were deemed satisfactory for the subsequent stages. The distribution of image captures for each camera, as well as the resulting calibration matrices and distortion parameters for each camera, can be found in Table IV.1 and IV.2 below.

## Stereo Calibration for Multi-Camera Setup

In this experiment, the process involves capturing calibration pattern images for the three cameras simultaneously, if feasible. Alternatively, the images are taken for one camera pair first and then for the other pair if simultaneous capture is not possible. This is due to the dependency on camera positions and rotations, which might hinder all three cameras from capturing calibration patterns concurrently.

The obtained rotation and translation results from the stereo calibration process for each test scenario are presented in Table IV.2 for subsequent analysis.

It is crucial to note that the rotations and translations of the cameras are measured in terms of the angular differences and displacements relative to the main camera, i.e., the center camera. The angular difference and displacement of the center camera itself are set to zero, as it remains unaltered concerning its own position and orientation.

## Evaluation of 3D Estimation Results: Triangulation and Bundle Adjustment

In this experiment, we compare the performance of 3D estimation results obtained through triangulation and those optimized with bundle adjustment. To facilitate the comparison, the discussion and analysis of both processes are presented in this subchapter.

To achieve the best triangulation results, we conducted tests for four rotation ranges between the cameras: 0°-10°, 25°-35°, 40°-50°, and 51°-70°, as discussed in subchapter III.3.5.

For each test range, we identified the test number and the camera pair that resulted in the lowest reprojection error. Data on the 2D marker locations during the actor's upright standing pose were provided for each test, as shown in Figure IV.5.

The reprojection error results for each test are presented in Table IV.3.

Based on the findings from Table IV.3, the following observations were made:

There is no significant impact of camera rotation changes on the reprojection error. This confirms the accuracy of the calibration process for both intrinsic and extrinsic camera matrices as well as the usage of 2D data.

The application of bundle adjustment significantly influences the reduction of mean reprojection error. This confirms the successful and beneficial application of bundle adjustment for optimizing 3D estimation.

Prior to bundle adjustment, triangulation using all three cameras simultaneously resulted in the lowest mean reprojection error. This suggests that the use of more than two cameras or multi-camera setups can reduce reprojection error.

After bundle adjustment, triangulation using all three cameras simultaneously did not always result in the lowest mean reprojection error. This could be due to the better initial estimates for the triangulation pairs, making it easier for the optimization algorithm to find the optimal solution. The optimization algorithm relies on derivative principles and is sensitive to the quality of initial estimates, leading to varied results.

Further evaluations of the 3D quality are conducted, focusing on the overall 3D structure rather than the reprojection error, which is qualitative rather than quantitative. Three motion tests—pose-A (raising and lowering the actor beside the body), left-right movement, and forward-backward movement—are performed to assess the system's ability to estimate simple to complex movements. These tests are conducted for the predetermined four test ranges.

Key findings from the experiments are as follows:

Overall, most movements are accurately captured. However, some 3D structure results show discrepancies or inaccuracies in depth estimation. For instance, Figure IV.4 shows a "floating" 3D structure, indicating depth estimation errors. The analysis revealed that asynchrony between the camera frames caused erroneous pairings of 2D points, leading to depth estimation errors. This demonstrates that the use of three cameras does not always guarantee the best results, as each camera might detect different points.

The influence of bundle adjustment on 3D estimation results is not easily noticeable. Minor improvements, approximately 2 pixels, were observed in the 3D structure, which were not visually apparent.

Reprojection error serves as an indication of inaccurate 3D estimation. As shown in Figure IV.7, when the 3D results "float," the reprojection error increases, as indicated by the yellow-colored areas in the bottom panel. This correlation between reprojection error and 3D results highlights the significance of reprojection error as an indicator of the quality of the obtained 3D structure, where incorrect 3D estimations fail to align with their corresponding 2D data points.

## Units

* Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as “3.5-inch disk drive.”
* Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
* Do not mix complete spellings and abbreviations of units: “Wb/m2” or “webers per square meter,” not “webers/m2.” Spell units when they appear in text: “...a few henries,” not “...a few H.”

Identify applicable sponsor/s here. If no sponsors, delete this text box (*sponsors).*

* Use a zero before decimal points: “0.25,” not “.25.” Use “cm3,” not “cc.” (*bullet list*)

## Equations

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

Number equations consecutively. Equation numbers, within parentheses, are to position flush right, as in (1), using a right tab stop. To make your equations more compact, you may use the solidus ( / ), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in

*a**b*    

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use “(1),” not “Eq. (1)” or “equation (1),” except at the beginning of a sentence: “Equation (1) is ...”

## Some Common Mistakes

* The word “data” is plural, not singular.
* The subscript for the permeability of vacuum **0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter “o.”
* In American English, commas, semi-/colons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
* A graph within a graph is an “inset,” not an “insert.” The word alternatively is preferred to the word “alternately” (unless you really mean something that alternates).
* Do not use the word “essentially” to mean “approximately” or “effectively.”
* In your paper title, if the words “that uses” can accurately replace the word using, capitalize the “u”; if not, keep using lower-cased.
* Be aware of the different meanings of the homophones “affect” and “effect,” “complement” and “compliment,” “discreet” and “discrete,” “principal” and “principle.”
* Do not confuse “imply” and “infer.”
* The prefix “non” is not a word; it should be joined to the word it modifies, usually without a hyphen.
* There is no period after the “et” in the Latin abbreviation “et al.”
* The abbreviation “i.e.” means “that is,” and the abbreviation “e.g.” means “for example.”

An excellent style manual for science writers is [7].

# Using the Template

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

## Authors and Affiliations

The template is designed so that author affiliations are not repeated each time for multiple authors of the same affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization). This template was designed for two affiliations.

### For author/s of only one affiliation (Heading 3): To change the default, adjust the template as follows.

#### Selection (Heading 4): Highlight all author and affiliation lines.

#### Change number of columns: Select the Columns icon from the MS Word Standard toolbar and then select “1 Column” from the selection palette.

#### Deletion: Delete the author and affiliation lines for the second affiliation.

### For author/s of more than two affiliations: To change the default, adjust the template as follows.

#### Selection: Highlight all author and affiliation lines.

#### Change number of columns: Select the “Columns” icon from the MS Word Standard toolbar and then select “1 Column” from the selection palette.

#### Highlight author and affiliation lines of affiliation 1 and copy this selection.

#### Formatting: Insert one hard return immediately after the last character of the last affiliation line. Then paste down the copy of affiliation 1. Repeat as necessary for each additional affiliation.

#### Reassign number of columns: Place your cursor to the right of the last character of the last affiliation line of an even numbered affiliation (e.g., if there are five affiliations, place your cursor at end of fourth affiliation). Drag the cursor up to highlight all of the above author and affiliation lines. Go to Column icon and select “2 Columns”. If you have an odd number of affiliations, the final affiliation will be centered on the page; all previous will be in two columns.

## Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include ACKNOWLEDGMENTS and REFERENCES, and for these, the correct style to use is “Heading 5.” Use “figure caption” for your Figure captions, and “table head” for your table title. Run-in heads, such as “Abstract,” will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles named “Heading 1,” “Heading 2,” “Heading 3,” and “Heading 4” are prescribed.

## Figures and Tables

### Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 1,” even at the beginning of a sentence.

1. Table Styles

| Table Head | Table Column Head | | |
| --- | --- | --- | --- |
| Table column subhead | Subhead | Subhead |
| copy | More table copya |  |  |

1. Sample of a Table footnote. *(Table footnote)*
2. Example of a figure caption. *(figure caption)*

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization,” or “Magnetization, M,” not just “M.” If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization (A ( m(1),” not just “A/m.” Do not label axes with a ratio of quantities and units. For example, write “Temperature (K),” not “Temperature/K.”

##### Acknowledgment *(Heading 5)*

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g.” Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

##### References

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

1. G. Eason, B. Noble, and I.N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955. (*references*)

We suggest that you use a text box to insert a graphic (which is ideally a 300 dpi resolution TIFF or EPS file with all fonts embedded) because this method is somewhat more stable than directly inserting a picture.

To have non-visible rules on your frame, use the MSWord “Format” pull-down menu, select Text Box > Colors and Lines to choose No Fill and No Line.

1. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.
2. I.S. Jacobs and C.P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.
3. K. Elissa, “Title of paper if known,” unpublished.
4. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
5. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
6. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.