Báo cáo môn Xử lý ảnh và thị giác Robot: Xác định vị trí của vật thể 3D dựa trên phương pháp hồi quy trong hệ thống thị giác máy

Lê Anh Chiến

Tháng 12 năm 2023

Tóm tắt nội dung

Thị giác máy hay thị giác robot đóng vai trò ngày càng quan trọng trong nhiều hệ thống công nghiệp, hứa hẹn tiềm năng rộng rãi trong tương lai của tự động hóa, bao gồm quản lý robot nội bộ, điều khiển robot theo đàn, giám sát dây chuyền sản xuất và thao tác gắp robot. Một trong những nhiệm vụ cơ bản nhưng đầy thách thức của thị giác máy là xác định vị trí 3D của vật thể. Dù đã có nhiều nghiên cứu đạt được kết quả tốt, việc xác định vị trí vật thể chính xác vẫn còn nhiều phần để cải thiện. Trong báo cáo này, tôi trình bày một thuật toán xác định vị trí vật thể 3D. Thuật toán sử dụng phương pháp dựa trên mẫu bàn cờ để thực hiện hiệu chuẩn camera, tiếp theo dựa trên một mô hình học sâu để xác định vị trí vật thể trong ảnh, sau đó nội suy vị trí tọa độ của vật thể thực tế và cuối cùng tinh chỉnh độ chính xác bằng mô hình hồi quy.

1 Giới thiệu

Thị giác máy là lĩnh vực nghiên cứu về việc xử lý và hiểu thông tin từ hình ảnh. Thị giác máy đóng vai trò ngày càng quan trọng trong nhiều lĩnh vực, bao gồm tự động hóa, robot, y tế và an ninh. Một trong những nhiệm vụ cơ bản của thị giác máy là xác định vị trí của vật thể 3D trong không gian. Xác định vật thể 3D có nhiều ứng dụng quan trọng như quản lý robot nội bộ, điều khiển robot theo đàn, giám sát dây chuyền sản xuất, thao tác gắp robot. Trong quản lý robot nội bộ, robot cần xác định vị trí của các vật thể trong nhà máy để thực hiện các nhiệm vụ như vận chuyển, lắp ráp và sửa chữa. Với điều khiển robot theo đàn thì mỗi robot cần xác định vị trí của các robot khác nhau trong đàn để phối hợp hoạt động. Trong giám sát dây chuyền sản xuất, hệ thống giám sát cần xác định vị trí của các sản phẩm trên dây truyền để kiểm tra chất lượng và hiệu suất. Và ứng dụng phổ biến nhất là thao tác gắp của robot thì robot xác định vị trí và kích thước của vật thể cần gắp để thực hiện thao tác gắp chính xác.

Có nhiều phương pháp khác nhau để xác định vị trí vật thể 3D. Các phương pháp này có thể được phân loại thành hai nhóm chính bao gồm phương pháp trực tiếp và phương pháp gián tiếp. Đối với phương pháp trực tiếp, các đặc điểm của vật thể được sử dụng để xác định vị trí 3D của nó. Còn với phương pháp gián tiếp thì các thông tin từ môi trường xung quanh được sử dụng để xác định vị trị của vật thể.

Trong báo cáo này, tôi giới thiệu một thuật toán xác định vật thể 3D dựa trên phương pháp hồi quy. Thuật toán này sử dụng một camera 2D đơn giản để thu thập dữ liệu hình ảnh. Thuật toán gồm bốn bước chính bao gồm hiệu chuẩn camera, xác định tâm vật thể trong hệ tọa độ ảnh, nội suy tọa độ của vật thể trong hệ tọa độ thực và cuối cùng là sử dụng một mô hình hồi quy để tinh chỉnh lại vị trí của vật thể.

Phần còn lại của báo cáo này gồm 3 phần. Phần 2 trình bày các nghiên cứu liên quan đến việc xác định vị trí vật thể và thực hiện hiệu chuẩn camera. Phần 3 trình bày các thiết lập về hệ thống và phương pháp được sử dụng. Phần 4 trình bày về các kết quả đạt được cũng như các đánh giá dựa trên các kết quả đã đạt được.

2 Các nghiên cứu liên quan

Trong các hệ thống thị giác máy, quá trình hiệu chuẩn camera là một bước quan trọng để lấy thông tin từ hình ảnh 2D để hiểu vật thể 3D thực tế và để xác định tỷ lệ pixel/mm giữa vật thể được chiếu trong ảnh và vật thể 3D thực. Thông tin này là cơ bản để đánh giá chính xác đối tượng đang được kiểm tra hoặc chọn hàng. Để đạt được sự định vị đối tượng, một số phương pháp đã được giới thiệu như phương pháp dây chuẩn [1] và phương pháp hai giai đoạn [2].

Phương pháp dây chuẩn [1] là một trong số ít các phương pháp hiệu chuẩn camera sử dụng mô hình thực tế để giải quyết vấn đề lệch tâm ống kính do sai lệch trong sản xuất. Tuy nhiên, phương pháp này có nhược điểm là phải xác định thủ công các điểm hiệu chuẩn và khó khăn trong việc xác định điểm chính thực tại vị trí lệch khỏi trục lý tưởng của ống kính. Phương pháp hai giai đoạn [2] của Tsai sử dụng cùng mô hình camera với [1], nhưng tập trung hơn vào đặc điểm vận hành theo thời gian thực. Phương pháp này sử dụng mẫu bàn cờ để tính toán hệ số tỉ lệ trong giai đoạn đầu và tính toán hiệu quả của tiêu cự, hệ số biến dạng trong giai đoạn thứ hai. Do giả định về một mô hình camera đơn giản (ví dụ: mô hình lỗ kim) và bỏ qua sự chiếu của vật thể 3D lên ảnh 2D, phương pháp này vẫn có hạn chế trong nhiều trường hợp.

Các phương pháp được đề xuất trong [3]-[4] dựa trên giả định biến đổi tuyến tính trực tiếp (DLT). Chiến lược là tìm các hệ số biến dạng xuyên tâm và tiếp tuyến chỉ dựa trên biến đổi tuyến tính. Các nhà nghiên cứu [5], [6] sử dụng DLT để đơn giản hóa thuật toán của phương pháp dây chuẩn. Sturm và Maybank [7] đề xuất sử dụng mẫu hiệu chuẩn 3D gồm 3 mặt phẳng với các chấm được hiệu chuẩn trên mỗi mặt phẳng. Phương pháp này đơn giản để thực hiện và tổng quát, nhưng đòi hỏi một mẫu hiệu chuẩn 3D chính xác.

Trong khi đó, nghiên cứu của Zhang [8], [9] đề xuất một kỹ thuật hiệu chuẩn camera khác. Phương pháp này chỉ yêu cầu một mẫu phẳng đơn giản nhưng cần chụp nhiều ảnh của mẫu được chụp từ các hướng khác nhau, điều này không chắc chắn và cũng khó triển khai trong bối cảnh công nghiệp.

Các phương pháp hiện đại [10] - [13] có xu hướng sử dụng học sâu để ước tính hiệu quả các tham số trong mô hình camera [10], [11] hoặc sử dụng các thiết bị đắt tiền (ví dụ: lidar hoặc camera 3D) để định vị vật thể trên trục z [12] -[13]. Những phương pháp này hoặc chỉ tập trung vào việc hoàn nguyên hình ảnh và không xem xét vị trí thực tế của từng vật thể trên ảnh (do giả thuyết học sâu) hoặc yêu cầu phần mềm hiện đại, tốn kém và không phù hợp trong bối cảnh công nghiệp.

Bài báo cáo này được thực hiện dựa trên ý tưởng sử dụng một mô hình hồi quy [14] để hiệu chỉnh lại vị trí của vật thể sau khi nội suy tâm vật bằng các ma trận thu được sau khi thực hiện hiệu chuẩn camera.

3 Hệ thống và phương pháp

3.1 Thiết lập hệ thống

Để thực hiện bài toán xác định vị trí của vật thể 3D, tôi thiết lập một hệ thống đơn giản với 3 thành phần chính bao gồm một mẫu bàn cờ, một camera và hai vật thể được sử dụng. Mẫu bàn cờ tôi sử dụng là một mẫu bàn cờ vua có kích thước 3.3x3.3 cm mỗi ô và có kích thước là 8x8 ô như trong Hình 1(a). Camera được sử dụng là camera của điện thoại Realme 6i được đặt chính giữa mẫu bàn cờ và cách mặt bàn cờ một khoảng cách cố định là 33 cm như Hình 1(b). Các vật thể được sử dụng là hai khối hình hộp có kích thước là 3.5x3.5x2.2 cm (Hình 1(c))và 3.5x7.0x2.2 cm (Hình 1(d)) được đặt tại các vị trí trên ô bàn cờ khác nhau để lấy tọa độ chính xác làm tiêu chí đánh giá độ chính xác.

3.2 Phương pháp

Trong báo cáo này, phương pháp tôi sử dựng được chia thành 4 bước bao gồm hiệu chuẩn camera, xác định tọa độ tâm vật trên hệ tọa độ ảnh, nội suy tọa độ tâm vật trên hệ tọa độ thực tế và cuối cùng là sử dụng phương pháp hồi quy để tinh chỉnh lại tọa độ của vật thể.

3.2.1 Hiệu chuẩn camera

Hiệu chuẩn camera là quá trình tìm các ma trận trong và ma trận ngoài (ma trận xoay và vector chuyển) của camera dựa trên một tập các điểm 3D đã biết theo tọa độ thực và tọa độ của các điểm đó tương ứng trên hệ tọa độ ảnh. Các ma trận trong và ma trận ngoài của camera giúp chuyển hệ tọa độ

(a) Mẫu bàn cờ

(b) Camera điện thoại

(c) Vật thể thứ nhất

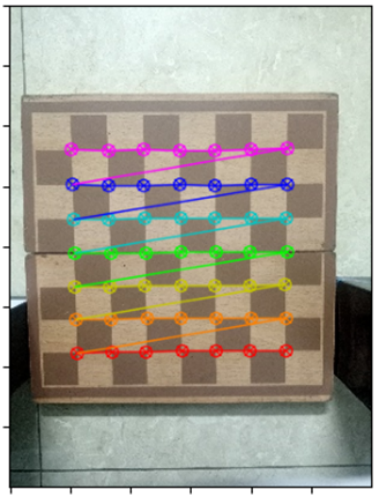
(d) Vật thể thứ hai

Hình 1: Các thành phần hệ thống

của một điểm trên ảnh sáng hệ tọa độ thực và ngược lại. Trong báo cáo lần này, tôi sử dụng phương pháp hiệu chuẩn bằng mẫu bàn cờ với các hàm có sẵn trong thư viện OpenCV. Quá trình hiệu chuẩn được chia làm 2 bước chính bao gồm sử dụng hàm cv2.findChessboardCorners() để tìm tọa độ (u, v) của các điểm 3D trên hệ tọa độ ảnh như Hình 2 và sử dụng hàm cv2.calibrateCamera() để xác định các ma trận xoay, vector chuyển và ma trận trong của camera. Các ma trận tìm được có các giá trị như sau:

$$\begin{aligned}
\mathbf{R} &= \begin{bmatrix} -0.9998768 & -0.00605625 & -0.01448117 \\ 0.00730226 & -0.99612896 & -0.08760009 \\ -0.01389459 & -0.08769504 & 0.99605046 \end{bmatrix}, \\
\mathbf{t} &= \begin{bmatrix} 0.28062436 \\ 0.40807267 \\ 1.8878581 \end{bmatrix}, \\
\mathbf{K} &= \begin{bmatrix} 535.02337469 & 0 & 147.67452617 \\ 0 & 516.43834232 & 174.1664992 \\ 0 & 0 & 1 \end{bmatrix},
\end{aligned} \tag{1}$$

trong đó $\mathbf{R}$ là ma trận xoay, $\mathbf{t}$ là ma trận chuyển và $\mathbf{K}$ là ma trận trong của camera.



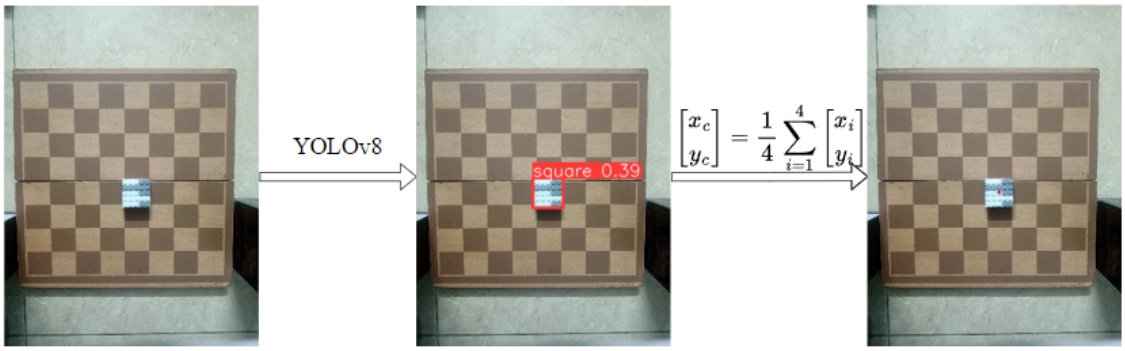
Hình 2: Minh họa khi sử dụng hàm cv2.findChessboardCorners

3.2.2 Xác định tọa độ tâm vật thể trên hệ tọa độ ảnh

Để xác định tâm vật thể trên hệ tọa độ ảnh, tôi sử dụng thuật toán YOLOv8 để xác định khung hình bao quanh vật thể và xác định được tọa độ của bốn điểm bao quanh vật thể. Tọa độ tâm của vật thể trên hệ tọa độ ảnh được xác định theo công thức sau:

$$\begin{bmatrix} x_c \\ y_c \end{bmatrix} = \frac{1}{4} \sum_{i=1}^{4} \begin{bmatrix} x_i \\ y_i \end{bmatrix}, \tag{2}$$

trong đó, $\begin{bmatrix} x_c \\ y_c \end{bmatrix}$ là tọa độ tâm của vật thể, $\begin{bmatrix} x_i \\ y_i \end{bmatrix}$ là tọa độ điểm thứ i bao quanh vật thể. Quá trình thực hiện xác định tọa độ tâm vật trên hệ tọa độ ảnh được trực quan hóa như Hình 3.



Hình 3: Ví dụ về xác định tâm vật trên hệ tọa độ ảnh

3.2.3 Xác định tọa độ tâm vật thể trên hệ tọa độ thực

Sau khi xác định được tọa độ của tâm vật trên hệ tọa độ ảnh và các ma trận của camera thông qua quá trình hiệu chuẩn thì tọa độ tâm của vật thể trên hệ tọa độ thực được nội suy từ các thông tin trên theo công thức như sau :

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{R}^{-1} \left(\mathbf{K}^{-1} \begin{bmatrix} u\mathbf{t}_1 \\ v\mathbf{t}_2 \\ \mathbf{t}_3 \end{bmatrix} - \mathbf{t} \right), \tag{3}$$

trong đó $[x, y, z]^T$ là tọa độ tâm vật thể trong hệ tọa độ thực, $[u, v]^T$ là tọa độ tâm vật thể trong hệ tọa độ ảnh.

3.2.4 Hiệu chỉnh tạo độ bằng phương pháp hồi quy

Để hiệu chinh lại tọa độ do các sai lệch gây ra bơir hình dạng vật hay góc chiếu của nguồn sáng, vị trí của vật thể so với camera, các tác giả trong bài báo [14] đã giới thiệu một phương pháp hiệu chỉnh bằng phương pháp hồi quy giúp giảm thiểu sai số giữa tâm vật thể được xác định trên ảnh và tâm vật thể thực tế. Đầu tiên, ta có mỗi quan hệ giữa khoảng cách của điểm hội tụ C với tâm vật thể thực tế và khoảng cách giữa tâm vật thể trên ảnh với tâm vật thể thực tế như (4).

$$||C - A(i)||_2 = \alpha ||P(i) - A(i)||_2 + \beta \tag{4}$$

Trong đó, vị trí của điểm C là cố định, $A(i)$ và $P(i)$ là tọa độ tâm vật thể thực tế và tâm vật thể trên ảnh của vật thể mẫu thứ i. α, β là các tham số của mô hình. Tham số β có thể được thêm để điều chỉnh sai số gây ra bởi hình dạng vật thể, góc nghiêng camera, ... Trong bài báo thì các tác giả sử dụng $\beta = 0$. Vị trí của điểm hội tụ $C(x_C, y_C)$ và α, β sử dụng hồi quy tuyến tính với tối ưu hóa bình phương nhỏ nhất, tức là cực tiểu hóa hàm mục tiêu như (5).

$$\alpha, \beta, x_C, y_C = \arg\min f(\alpha, \beta, x_C, y_C) \tag{5}$$

với

$$f(\alpha, \beta, x_C, y_C) = \frac{1}{n}\sum_{i=1}^{N}\left(\alpha - \frac{\sqrt{(x_C - x_A(i))^2} - \beta}{\sqrt{(x_P(i) - x_A(i))^2 + (y_P(i) - x_A(i))^2}}\right)^2 \tag{6}$$

Do $\beta = 0$ nên α được tính theo công thức như (7)

$$\alpha = \sum_{i=1}^{n}\frac{\sqrt{(x_C - x_A(i))^2 + (y_C - y_A(i))^2}}{\sqrt{(x_P(i) - x_A(i))^2 + (y_P(i) - x_A(i))^2}} \tag{7}$$

Như vậy, (x_C, y_C) sẽ được tính thông thông qua một thuật toán gradient descent để ước lượng nghiệm của (6) như thuật toán 1. Cuối cùng, sau khi thu được các tham số của mô hình hồi quy thì α, β và

Algorithm 1 Ước lượng điểm hội tụ bằng gradient desent

1: Khởi tạo ngẫu nhiên tập $(x_C(0), y_C(0)$. Tính toán α_0 theo (7).
2: Cập nhật giá trị của $(x_C(k+1), y_C(k+1)$ theo $(x_C(k), y_C(k)$ như sau:

$$x_C(k+1) = x_C(k) - \sigma\frac{df}{dx_C(k)} \tag{8}$$

$$y_C(k+1) = y_C(k) - \sigma\frac{df}{dy_C(k)} \tag{9}$$

trong đó, σ là tốc độ học và được đặt $\sigma = 0.01$.
3: Tính toán $\alpha(k+1)$ theo (7).
4: Nếu $\frac{df}{dx_C(k)}$ và $\frac{df}{dy_C(k)}$ là đủ nhỏ thì dừng thuật toán. Nếu không tiếp tục lại từ bước 2.

tọa độ điểm hội tụ (x_C, y_C) được đưa vào phương trình (4) để đưa ra vị trí thực tế của vật thể.

4 Kết quả và đánh giá

Trong báo cáo lần này, tôi thực hiện quá trình huấn luyện mô hình hồi quy với 15 ảnh mẫu đầu vào và thực hiện kiểm thử trên 20 ảnh trong bộ dữ liệu. Tôi thực thi phương pháp đã trình bày trên ngôn ngữ lập trình Python, các thư viện hỗ trợ như Numpy, Pytorch, OpenCV và YOLOv8. Bảng 4 và Hình 4 mô tả sai số vị trí khi xác định tâm của vật thể dự đoán được với vị trí thật của vật thể dựa trên trục x và trục y và khoảng cách Euclid ($\Delta = \sqrt{\Delta x^2 + \Delta y^2}$ khi áp dụng phương pháp đã được trình bày phần trên. Từ kết quả thu được, phương pháp đưa ra đã dự đoán được tọa độ tâm của vật thể 3D với sai số trung bình chỉ 2.27 mm và sai số nhỏ nhất chỉ khoảng 0.23 mm từ đó tăng khả năng hoàn thành các tác vụ trong việc sử dụng robot để gắp vật thể.

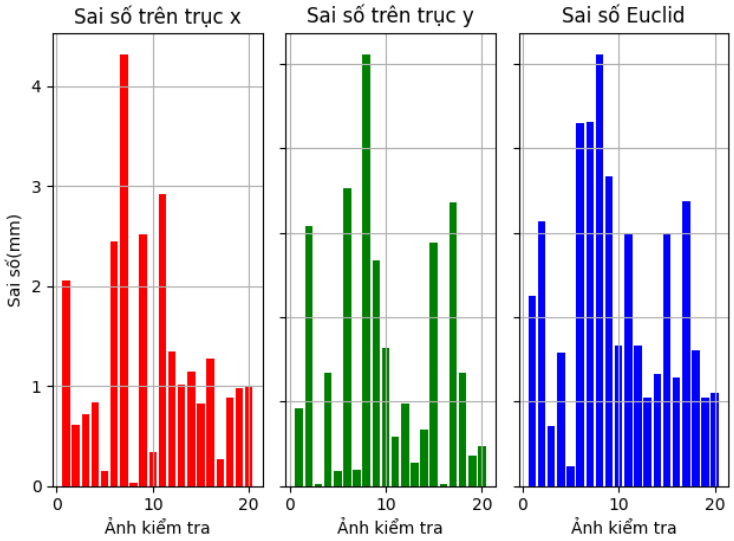
5 Kết luận

Trong báo cáo này, tôi đã trình bày một phương pháp xác định vị trí tâm vật thể 3D, một bước quan trọng và phổ biến trong các vấn đề học máy. Phương pháp đã trình bày được tạo dựa trên mối quan hệ hình học giữa vị trí thực của vật thể và thông tin chiếu của nó thu được bằng thuật toán YOLO. Phương pháp đã trình bày có sai số trung trung bình của vị trí vật thể khoảng 2.27 mm, đủ nhỏ để triển khai trong hệ thống robot gắp. Đối với công việc trong tương lai, tôi có thể cải thiện độ chính xác của mô hình hơn nữa với quá trình học trực tuyến.

References

[1] C. B. Duane, "Close-range camera calibration," *Photogramm. Eng*, vol. 37, no. 8, pp. 855–866, 1971.

[2] R. Tsai, "A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses," *IEEE Journal on Robotics and Automation*, vol. 3, no. 4, pp. 323–344, 1987.



Hình 4: Kết quả đánh giá sai số

Bảng 1: Kết quả đánh giá sai số

Mẫu	Δx	Δy	Δ
1	2.06	0.92	2.25
2	0.62	3.08	3.14
3	0.71	0.02	0.71
4	0.83	1.34	1.58
5	0.15	0.18	0.23
6	2.45	3.53	4.30
7	4.32	0.19	4.32
8	0.03	5.11	5.11
9	2.52	2.67	3.67
10	0.34	1.64	1.67
11	2.92	0.58	2.98
12	1.34	0.98	1.66
13	1.02	0.28	1.05
14	1.14	0.68	1.32
15	0.83	2.88	3.00
16	1.28	0.03	1.28
17	0.27	3.36	3.37
18	0.88	1.34	1.60
19	0.98	0.36	1.04
20	1.00	0.47	1.11
Trung bình	1.28	1.48	2.27

[3] Y. I. Abdel-Aziz, H. M. Karara, and M. Hauck, "Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry," *Photogrammetric engineering & remote sensing*, vol. 81, no. 2, pp. 103–107, 2015.

[4] J. Heikkila and O. Silvén, "A four-step camera calibration procedure with implicit image correction," in *Proceedings of IEEE computer society conference on computer vision and pattern recognition*, IEEE, 1997, pp. 1106–1112.

[5] T. A. Clarke and J. G. Fryer, "The development of camera calibration methods and models," *The Photogrammetric Record*, vol. 16, no. 91, pp. 51–66, 1998.

[6] J. Fryer, T. Clarke, and J. Chen, "Lens distortion for simple c-mount lenses," *International Archives of Photogrammetry and remote sensing*, vol. 30, pp. 97–101, 1994.

[7] P. F. Sturm and S. J. Maybank, "On plane-based camera calibration: A general algorithm, singularities, applications," in *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, IEEE, vol. 1, 1999, pp. 432–437.

[8] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *Proceedings of the seventh ieee international conference on computer vision*, Ieee, vol. 1, 1999, pp. 666–673.

[9] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[10] O. Bogdan, V. Eckstein, F. Rameau, and J.-C. Bazin, "Deepcalib: A deep learning approach for automatic intrinsic calibration of wide field-of-view cameras," in *Proceedings of the 15th ACM SIGGRAPH European Conference on Visual Media Production*, 2018, pp. 1–10.

[11] G. Iyer, R. K. Ram, J. K. Murthy, and K. M. Krishna, "Calibnet: Geometrically supervised extrinsic calibration using 3d spatial transformer networks," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2018, pp. 1110–1117.

[12] P. An, T. Ma, K. Yu, *et al.*, "Geometric calibration for lidar-camera system fusing 3d-2d and 3d-3d point correspondences," *Optics express*, vol. 28, no. 2, pp. 2122–2141, 2020.

[13] X. Pan, Z. Xia, S. Song, L. E. Li, and G. Huang, "3d object detection with pointformer," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 7463–7472.

[14] X. H. Van and N. Do, "An efficient regression method for 3d object localization in machine vision systems," *IAES International Journal of Robotics and Automation*, vol. 11, no. 2, p. 111, 2022.