UAV ranging method based on monocular vision

Yicheng Rong¹, Zhiyong Zhang¹*, Changzhen Qiu¹, Zezheng Ye¹

¹School of Electronics and Communication Engineering, Sun Yat-Sen University, Guangzhou, Guangdong, China

*Corresponding author’s e-mail: zhangzhy99@sysu.edu.cn

Abstract. Aiming at the problem that the unmanned aerial vehicle (UAV) cannot complete the positioning and ranging in the complex background and the signal is interfered, we proposed a monocular measurement scheme of assigning 4 LED fiducials (also called landmarks) with 3 colors in total on the target UAV. Firstly, feature areas are screened out by using high saturation, high brightness and specific hue range of LED in HSV (Hue, Saturation, Value) color space. After morphological processing, the background that does not meet the features is removed. Then, the mapping relationship between target LED fiducials and image fiducials is established by using the average hue of connected domain and special geometric layout. The improved Lambda-Twist is used to find the distance from the target UAV to the center of the camera, and the performance of the UAV recognition and ranging from 90m to 245m is tested. The experimental results show that the method can quickly and accurately measure the distance to the target, and can be used for UAV auxiliary measurement.

1. Introduction
The distance relationship between unmanned aerial vehicles (UAVs) is an important condition for formation flight to maintain relative stability. Global Position System (GPS) positioning is a technology widely used in UAV positioning at present. In areas with weak GPS signals or signal interference, other means are needed to assist in measuring distance information. The commonly used ranging methods include laser ranging, radar ranging and visual ranging. Vision equipment is widely used in auxiliary measurement due to its greater advantages in volume, weight, price and power consumption on UAV platform. Because binocular imaging requires a long baseline when measuring long-range targets[1], it is quite difficult to arrange in UAV.

Monocular imaging has the advantages of fast processing speed, simple structure and high positioning accuracy, which meet the requirements of UAV-aided ranging. The distance measurement based on monocular lens can be divided into two steps. Firstly, image information is obtained through image processing, and then the attitude distance is solved by using the information. Feature point extraction is the primary premise of distance measurement. Common feature point extraction algorithms include Harris corner point[2], GFTT corner point[3], FAST corner point[4] and other direct feature extraction algorithms, as well as local image features such as SIFT[5], ORB[6], SURF[7]. There are also recently popular neural networks to obtain feature point information, such as R-CNN series[8] of two-stage method and single-stage YOLO series[9, 10], but it is difficult to give consideration to both real-time performance and accuracy. The artificial reference point is easier to extract and more stable to detect, and it is also of great help to solve the attitude[11]. Such as the reference point of concentric circles type, namely put black rings on a white background, or conversely, has the very good effect. The center of mass of the circular remained relatively unchanged under the perspective of deformation, and...
it can be easily positioned with sub-pixel accuracy[12], but it is not applicable under different lighting conditions, and the applicability of the five-point geometry layout proposed by it is also harsh. By changing the color of the rings to other color and the special distribution of the rings, the reference point can be applied under different light conditions[13]. Due to its high brightness, light-emitting diode (LED) lights are often used as the signal reference point of the target. For example, Bajura arranged multiple monochrome LED lights on the target as signal signs[14], which requires very high accuracy of attitude estimation of the initial target.

In order to meet the purpose of rapid and accurate distance measurement of UAV under complex environmental conditions, this article is a special fiducial setting scheme with 3 colors and 4 lights. As few lights as possible are set at suitable positions on the UAV, the identification of the corresponding relationship is simplified. We use as few colors as possible to improve the efficiency of fiducials recognition, and use the improved Lambda-TWSIT[15] method to solve the attitude of this scene, which improve the recognition speed of the UAV target attitude distance. Finally, the video ranging of the UAV semi-physical simulation flight is given, and the corresponding results are compared and analyzed.

2. Extraction of three-color LED fiducials arranged on the UAV

The installation of easy-to-detect signal signs on the target UAV is still an important technical means for rapid identification and detection. Since it is difficult to place signs such as concentric circles on the surface of the UAV, and the detection of concentric circles is not suitable for dark night situations, this system chooses LED as the fiducials. It is easy to install and can work all day long, and it also has better anti-blocking ability for targets with large changes in attitude such as UAV. The extraction of fiducials in this system includes two steps: image fiducials detection, and the relationship between image fiducials and target fiducials.

2.1. Layout of LED lights

By carefully setting the fiducials, it can be easily detected and recognized, and the performance of attitude estimation can be effectively improved. At least three points are required to obtain a closed-form solution. P3P problems have up to four sets of positive solutions. Four control points are coplanar to determine a unique attitude solution[16]. For the stability of ranging, we install at least four LED lights. However, the fewer types of LED colors installed on the UAV, the more it can avoid the background color, reduce the difficulty of recognition, and the higher the robustness. The selection of a single color LED requires a more accurate estimation of the initial target attitude[14]. Based on the accuracy and convenience of post-processing considerations, we chose to use 3 colors and 4 lights. Red LED lights of the same color are placed on both ends of the UAV wing, and yellow and blue lights are placed on both ends of the tail wing, as shown in Figure 1, so that the four lights are in the same plane and separated as far as possible.

![Figure 1](image_url)

Figure 1. The distribution of 3 colors and 4 lights on the UAV.
2.2. Image LED fiducials detection

2.2.1. Image processing
In this system, the photoelectric pod of the rear UAV has roughly framed the position of the front UAV in the middle of the image. In order to reduce the amount of calculation and save calculation time, we only detect the position of the 400*400 window in the center of the image. And on this basis, the image is converted into HSV (Hue, Saturation, Value) space to complete the image preprocessing of the system.

For LED lights, the saturation is higher and the brightness is higher than the surrounding environment, and the HSV space can capture this information. In addition, in this color space, it can also be distinguished from the white body, which makes it easier to detect the LED fiducials. However, in other color spaces, the brightness value of the darker white and the high-brightness pure color are the same. It is not conducive to distinguish the white body from the color LED, which is also an important reason for choosing the HSV color space.

2.2.2. Color feature detection
Based on the high brightness and high saturation of pure color LEDs in HSV space, the threshold segmentation of red, yellow and blue colors and the high brightness and high saturation was carried out to obtain the characteristic distribution of specific colors.

Start by creating a 400*400 image sxv with all values initialized to 0. In detection, the product of saturation S and brightness V of each pixel of the window is calculated first, and the sxv is corresponding one by one. In the specific image, the color areas with high brightness and saturation are retained, and the white body areas with low brightness and saturation are screened out. This distinguishes the LED fiducials from the UAV, preserving only areas of high saturation and brightness.

Then we need to pick out the color we want, that is, limit the hue value range. Table 1 shows the corresponding relationship between hue value and color. It can be seen that red, yellow and blue are 0°, 60° and 240° respectively. Divide the color is not the absolute value of the color, should be divided into the desired color near the hue value.

| hue (°) | color |
|---------|-------|
| 0       | red   |
| 60      | yellow |
| 120     | green |
| 180     | cyan  |
| 240     | blue  |
| 300     | magenta |

Measuring the similarity to a fixed hue requires calculating the distance between the hue and the fixed hue, and the hue value has periodicity, for example, 5 degrees and 355 degrees should be 10 degrees apart instead of 350 degrees. So, the formula for the distance between the hue and the fixed hue is:

\[ d = \begin{cases} |h - h_m|, & \text{if } |h - h_m| < 180° \\ 360° - |h - h_m|, & \text{others} \end{cases} \quad (1) \]

Create a 400*400 image hue1, calculate the Gaussian function (2) for the hue value of each pixel in the window, where \( h_m = 0 \), and put the calculated values into hue1 one by one. This means that the closer to the hue value 0, the more complete the value will be retained. Similarly, create hue2 and hue3 images with \( h_m \) of 60° and 240° respectively. The values of hue1, hue2, and hue3 images are correspondingly added and put into a new 400*400 image mask, which is used to filter the tonal values.
Next, create a new 400*400 image bin. If the value of each pixel in the image sxv multiplied by the corresponding point in the image mask is greater than our preset threshold, the corresponding pixel in the bin is set to 255, otherwise it is set 0, the binary image bin is finally obtained, that is, the distribution image that meets the color expectation.

2.2.3. Screening image LED fiducials
The connected components are labeled on the binary graph bin, and the area, position, shape and other information of each connected domain are obtained, and the connected components whose area is too small and the shape is not very round are ignored. The shape of the filter is calculated by the following formula (3), (4), (5). When the ratio is less than 0.5 or greater than 2, this non-circular connected domain is excluded. Among them, CC_right, CC_left, CC_bottom, and CC_top are the rightmost ordinate, the leftmost ordinate, the bottom abscissa, and the uppermost abscissa of the connected domain, respectively.

\[
\text{width} = \text{CC}_{\text{right}} - \text{CC}_{\text{left}} + 1 \quad (3)
\]
\[
\text{height} = \text{CC}_{\text{bottom}} - \text{CC}_{\text{top}} + 1 \quad (4)
\]
\[
\text{ratio} = \frac{\text{width}}{\text{height}} \quad (5)
\]

Then find out all the points in each connected component, the mean value and standard deviation of the hue value of the points in the corresponding component, and exclude the connected domains with large standard deviations, that is, the non-LED areas with large color changes. Then find out all the points in each connected component, the mean value and standard deviation of the hue value of the points in the corresponding component, and exclude the connected domains with large standard deviations, that is, the non-LED areas with large color changes. The calculation of the mean and standard deviation requires special processing due to the periodicity of the hue value. Calculate the mean and standard deviation of the two intervals [-180°, 180°] and [0°, 360°] for each connected component. The smaller standard deviation of the two intervals is used as the statistical data of the connected component. For the hue value of [-180°, 0°] in the interval [-180°, 180°], it is converted from [180°, 360°]:

\[
h = \begin{cases} 
  h - 360°, & \text{if } h > 180° \\
  h, & \text{others}
\end{cases} \quad (6)
\]

2.3. Corresponding relationship between image fiducials and target fiducials
After getting the filtered connected components, we need to determine the correspondence between the image fiducials and the four LED lights. For yellow and blue LEDs, take the geometric center of the smallest distance between the hue value of the connected domain of 60 (yellow) and 240 (blue) as the corresponding target fiducials. Similarly, the red fiducials are determined by taking the geometric center closest to hue 0, but there are two red image fiducials, which cannot directly correspond to the red LEDs on both sides of the wings, and specific calculations are required to determine the corresponding relationship. As shown in Figure 2, we need to determine which is 1 (yellow light on the same side) and which is 2 (blue light on the same side). The line between the red lights 1 and 2 on the wing of the UAV is parallel to the line between the yellow lights 3 and blue lights 4 on the tail wing, so that the middle of the two points 1 and 2 points to the two points 3 and 4. The direction of the midpoint is \( \mathbf{v} \), the vector perpendicular to \( \mathbf{v} \) is \( \mathbf{v}^T \), the vector from 1 to 2 is \( \mathbf{v}_1 \), and the vector from 3 to 4 is \( \mathbf{v}_2 \). We found that the angle between \( \mathbf{v}_1 \) and \( \mathbf{v}^T \) and the angle between \( \mathbf{v}_2 \) and \( \mathbf{v}^T \) are always the same. However, in the perspective view, \( \mathbf{v}_1 \) and \( \mathbf{v}_2 \) are not completely parallel, but they are both acute or obtuse angles, and their inner products also have the same sign, so:

\[
\text{sign} = (\mathbf{v}^T \cdot \mathbf{v}_1) \cdot (\mathbf{v}^T \cdot \mathbf{v}_2) \quad (7)
\]

If sign>0, the order of the extracted image fiducials 1, 2 is just right. If sign<0, the order is reversed, and the coordinates of the two points are swapped. So far, we have found the image fiducials and corresponded with the target fiducials one-to-one.
3. Pose estimation

After obtaining the LED fiducials, Lambda-Twist can be used to solve the 3-point attitude, but the effect of direct use is not very ideal. For this scenario, we have made an adaptive improvement to it. The original will usually get two to four attitude solutions. In order to get the correct solution, the general algorithm is as follows: firstly, three points are used to calculate the attitude estimation, then the fourth target fiducial is used to calculate the projected position under each attitude estimation, and the error between this position and the correct position is calculated. The attitude solution with the smallest error is taken as the correct solution. Based on the actual attitude solution data, this algorithm finds that the correct attitude solution can be obtained by using only three of the four points through geometric determination.

3.1. Lambda-Twist algorithm principle

The equation for P3P to solve the attitude can be expressed as:

$$\lambda_i y_i = R x_i + t, i = \{1, 2, 3\}$$

(8)

where, $|y_i| = 1$, that is, $\lambda_i$ represents the absolute distance from the fiducial to the optical center of the camera. Eliminate the translation vector $t$ to obtain:

$$\lambda_i y_i - \lambda_j y_j = R(x_i - x_j)$$

(9)

$$|\lambda_i y_i - \lambda_j y_j|^2 = |x_i - x_j|^2$$

(10)

$$\lambda_i^2 + \lambda_j^2 - 2b_{ij}\lambda_i\lambda_j = a_{ij}$$

(11)

where, $a_{ij} = ||x_i - x_j||^2$, $b_{ij} = y_i^T y_j$, the solution is transformed into solve $A_k = (\lambda_1, \lambda_2, \lambda_3)$, then find the corresponding posture, as follows: let $z_1 = \lambda_1 y_1 - \lambda_2 y_2$, $z_2 = \lambda_2 y_2 - \lambda_3 y_3$, $Y = [z_1, z_2, z_1 \times z_2]$, $X = [x_1 - x_2, x_2 - x_3, (x_1 - x_2) \times (x_2 - x_3)]$, then:

$$R = YX^{-1}, t = \lambda_i y_i - Rx_i$$

(12) can be expressed in the form of a matrix as:

$$\begin{bmatrix} \lambda^T M_{12} A = a_{12}, \lambda^T M_{13} A = a_{13}, \lambda^T M_{23} A = a_{23} \end{bmatrix}$$

(13)

where

$$M_{12} = \begin{pmatrix} 1 & -b_{12} & 0 \\ -b_{12} & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, M_{13} = \begin{pmatrix} 1 & 0 & -b_{13} \\ 0 & 0 & 0 \\ -b_{13} & 0 & 1 \end{pmatrix}, M_{23} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & -b_{23} \\ 0 & -b_{23} & 1 \end{pmatrix}$$

(14)

Combine (13) linearly to get:

$$\begin{bmatrix} \lambda^T(a_{23} M_{12} - a_{12} M_{23}) A = A^T D_1 A = 0 \\ \lambda^T(a_{23} M_{13} - a_{13} M_{23}) A = A^T D_2 A = 0 \end{bmatrix}$$

(15)

(16)

then

$$\lambda^T(D_1 + \gamma D_2) A = 0$$

(17)
When \( det(D_1 + \gamma D_2) = 0 \), \( D_1 + \gamma D_2 \) corresponds to a pair of planes, which can be obtained by solving a third-degree polynomial, we set one of the solutions as \( \gamma_0 \). Then eigenvalue decomposition is performed for the singular matrix \( D_0 = D_1 + \gamma_0 D_2 \):

\[
D_0 = ESE^T = \begin{pmatrix}
\sigma_1 & 0 & 0 \\
0 & \sigma_2 & 0 \\
0 & 0 & 0
\end{pmatrix} E^T
\]  

(18)

where \( \sigma_1 > 0, \sigma_2 \leq 0 \), \( E = (e_1, e_2, e_3) \) is homography, then

\[
A^T D_0 A = A^T ESE^T A = 0
\]  

(19)

set \( p = (p_1, p_2, p_3)^T = E^T A \), then (19) is transformed as

\[
p^T S p = 0
\]  

(20)

and then

\[
p_1^2 \sigma_1 + p_2^2 \sigma_2 = 0, p_1 = sp_2, s = \pm \sqrt{-\frac{\sigma_2}{\sigma_1}}
\]  

(21)

\[
p_1 = sp_2 \rightarrow e_1^T A = s e_2^T A
\]  

(22)

\[
\lambda_1 = \frac{e_{21} - s e_{22}}{s e_{12} - e_{11}} \lambda_2 + \frac{e_{31} - s e_{32}}{s e_{12} - e_{11}} \lambda_3 = \omega_0 \lambda_2 + \omega_1 \lambda_3
\]  

(23)

set \( \lambda_3 = \tau \lambda_2 \), obviously \( \tau > 0 \), combine (23) and (15) to get

\[
\lambda_2^2 [\omega_0 + \tau \lambda_2, 1, \tau] D_1 [\omega_0 + \tau \lambda_2, 1, \tau]^T = 0
\]  

(24)

There are at most two solutions to this equation, and there are two cases for \( s \), that means, there are at most four \( \tau \) values, excluding negative and complex numbers. Substitute it into \( A^T M_{12} A = a_{12} \), we can get \( \lambda_2 = \sqrt{a_{23}/(b_{23} + \tau)} + 1 \), and then \( A = \lambda_2 [\omega_0 + \tau \omega_1, 1, \tau]^T \), and we can find the corresponding \( R \) and \( t \).

3.2. Improvements for this scenario

For the scene of this application, the main viewing angles are rear view and down view. The attitude solution obtained from this perspective is at most two solutions. The first case is the same solution, that means, both are correct solutions. The second case is two attitude solutions. The wrong solution can be eliminated by geometric judgment and the correct attitude solution can be obtained. The specific method is as follows:

The target coordinate system is shown in Figure 3 (take the midpoint of points 1, 2 as the origin O, the X axis is parallel to the wing, the Y axis points to the midpoint of points 3 and 4, and the Z axis is perpendicular to the XOY plane). In the postures corresponding to the two posture solutions, the position of the back camera in the Y-axis direction of the target coordinate system is one positive and the other negative. The rear UAV is always behind the front UAV, that is, the correct solution Y will be positive. Thus, when the Y coordinate of the camera in the target coordinate system is positive, that means, \( y = (0,0) \) is substituted into \( R^{-1}(y - t) = x \), and the Y coordinate is greater than 0, also the correct solution. Compared with the minimum projection error method, this improved method has a smaller amount of calculation and shorter calculation time, which is more valuable in this scenario.

At the same time, P3P’s long-distance solution will have a small chance of no solution. In this application, when there is no solution at 1, 2, and 3 points, we use 1, 2, and 4 to calculate the posture, which can greatly reduce the probability of no solution.
4. Experimental analysis

4.1. Experimental result

The article is based on videos of UAV’s semi-physical simulation flight, in which the length of UAV’s wing is 4.5m, the length of its tail is 1.495m, and the distance between the wing and the tail is 0.8372m. The internal parameters of the simulated camera are: pixel width 5.2*1000/1920um, focal length 4.1689*1000um, image resolution 1920*1080, running in Windows 10 system, Opencv 4.1.0, hardware intel i7-8700@3.20GHz, memory 16 GB. Use one of the frames to show the recognition effect. As shown in Figure 4, (a) is the 400*400 image intercepted in the video frame, (b) is the image sxv obtained by multiplying the S and V values of each pixel, (c) is the sxv_mask image obtained by correspondingly multiplying the sxv and the hue value screening matrix mask, (d) is the image bin obtained by the threshold segmentation, and (e) is the connected components distribution after excluding the non-LED color distribution law. Finally, the identified points are marked on the figure (f).

It can be seen that after the screening of saturation, brightness and hue values, the binary map bin has basically only left the LED lights area, and the connected components with too small and non-circular area is removed, namely (d). According to the average hue value and standard deviation, the geometric position of the four LEDs (connected components geometric center) is determined. Finally, according to geometric determination, the image fiducials correspond to the target fiducials. The experimental results show that the frame operation time is 5.3ms. After excluding the non-circular area and the area with too small area, the remaining 4 connected domains were marked out accurately. The distance between the camera and the target UAV was calculated to be 56.58m. The average time consumed by other frames is 5.2ms, which meets the requirements of high precision and fast calculation of this scene.
4.2. Algorithm performance analysis

In this application scenario, the expected range of measuring is 100–250m. In the actual algorithm, there will be no solution for the pose solution of three points. In this paper, when the improved Lambda-Twist algorithm has no solution for points 1, 2 and 3, the failure rate can be further reduced by using points 1, 2 and 4 to calculate again, as shown in Table 2. It can be seen that the failure rate can be reduced by using two combinations of four fiducials in remote scenarios above 220m. The main reason for the failure of
single operation solution is that the target distance is far away, and the accuracy of fiducial extraction and positioning is affected. There is a probability calculation failure, and the second combination of fiducials will also have a probability failure, but the probability of both failures is much smaller than the probability of single failure. Therefore, the success rate of distance ranging can be improved by calculating twice according to this rule:

Table.2  The effect of twice calculations on the success rate.

| Measuring range | Total video frames | Count of calculation | Number of no solution | Success rate  |
|-----------------|--------------------|----------------------|-----------------------|---------------|
| 90–156          | 1348               | once                 | 0                     | 100.00%       |
|                 |                    | twice                | 0                     | 100.00%       |
| 143–220         | 1485               | once                 | 0                     | 100.00%       |
|                 |                    | twice                | 0                     | 100.00%       |
| 188–245         | 1396               | once                 | 50                    | 96.42%        |
|                 |                    | twice                | 0                     | 100.00%       |

In the process of UAV ranging, real-time and accuracy are very important indicators. AP3P is one of the classical algorithms of recent P3P solution methods. It has very high numerical accuracy, but the algorithm is complex and takes a long time. Table 3 shows the comparison of the calculation time and error of the three algorithms based on 1348 frames of hardware-in-the-loop simulation video. It can be seen that the proposed algorithm has the same level of accuracy as AP3P algorithm in terms of error rate, but it takes much less time than AP3P. In addition, the improved algorithm in this paper is about 9% faster than the original Lambda-Twist solution by projection error. The reason is that, after solving the two attitude solutions, the improved algorithm only needs to calculate the Y coordinate of \( R^T (-t) \) to determine the correct attitude solution, while the original algorithm needs to substitute the fourth point into the two attitudes, and the error between the \( y_l = R_l x - t_l \) and the actual image projection point is the smallest, which is the correct solution. The improved algorithm has fewer calculation steps and the same calculation accuracy, so the improved algorithm is faster.

Table.3  Comparison of algorithm results based on 1348 frames of video.

| Algorithm               | Time cost/ms | Mean error rate | Maximum error rate |
|-------------------------|--------------|-----------------|--------------------|
| Algorithm used in this article | 181          | 0.34%           | 1.49%              |
| Original Lambda-Twist   | 199          | 0.34%           | 1.49%              |
| AP3P                    | 296          | 0.33%           | 1.48%              |

5. Conclusion
In this paper, based on the characteristics of high brightness and high saturation of LED in HSV color space, we proposed a method to measure the distance by carefully laying out 4 LED with 3 colors in total on the UAV. We use the threshold segmentation of high brightness, high saturation and specific hue value, as well as the morphological processing of the connected components to remove the area that is too small and non-circular. Then we use the average hue value and geometric relationship of the connected components to find the correspondence between image fiducials and target fiducials. After the fiducials are obtained, the improved Lambda-Twist method is used to solve the pose. The experimental comparative analysis shows that the improved algorithm reduces the probability of no solution under the long-distance target, and whose precision is equal to AP3P algorithm, but it is faster. Compared with the original Lambda-Twist algorithm, the speed is also slightly improved. The method can quickly detect the distance of the UAV and meet the requirements of real-time and accuracy as an auxiliary measurement under special circumstances.
References

[1] Zhao, X.J., Gao Z.Y. (2003) Photoelectric passive ranging technology. optical technology, 29(6): 652-656.
[2] Harris, C.G., Stephens, M.J. (1988) A combined corner and edge detector. In: Alvey vision conference.
[3] Shi, J., Tomasi, C. (1994) Good Features to Track. In: IEEE Computer-Society Conference on Computer Vision and Pattern Recognition. pp. 593-600.
[4] Rosten, E. (2006) Machine learning for high-speed corner detection. In: European Conference on Computer Vision.
[5] Lowe, D.G. (2004) Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, 60: 91-110.
[6] Rublee, E., Rabaud, V., Konolige, K., et al. (2011) ORB: an efficient alternative to SIFT or SURF. In: IEEE International Conference on Computer Vision. Barcelona. pp. 6-13.
[7] Bay, H. (2006) SURF: Speeded Up Robust Features. In: Computer Vision & Image Understanding. Berlin. pp. 110: 404-417.
[8] Ren, S., He, K., Girshick, R., et al. (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE transactions on pattern analysis and machine intelligence, 39:1137-1149.
[9] Redmon, J., Farhadi, A. (2018) YOLOv3: An Incremental Improvement. arXiv e-prints.
[10] Redmon, J., Divvala, S., Girshick, R., et al. (2015) You Only Look Once: Unified, Real-Time Object Detection.
[11] Hoff, W.A., Nguyen, K., Lyon, T. (2000) Computer-vision-based registration techniques for augmented reality. Proceedings of SPIE - The International Society for Optical Engineering, 2904:538-548.
[12] Lepetit, V., Fua, P. (2005) Monocular Model-Based 3D Tracking of Rigid Objects: A Survey. now, Hanover.
[13] Cho, Y.K. (1998) A-multi-ring color fiducial system and a rule-based detection method for scalable fiducial-tracking augmented reality. In: Proc. Int. Workshop on Augmented Reality (IWAR’98).
[14] Bajura, M., Neumann, U. (1995). Dynamic registration correction in video-based augmented reality systems: Computer Graphics and Applications, IEEE, 15, 52-60.
[15] Persson, M., Nordberg, K. (2018) Lambda Twist: An Accurate Fast Robust Perspective Three Point (P3P) Solver. In: European Conference on Computer Vision. Springer International Publishing, Cham. pp. 334-349.
[16] Fischler, M.A. (1981) A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 24: 726-740.