Straightforward Adaptation of Particle Filter to Fish Eye Images for Top View Pedestrian Tracking

Hicham Talaoubrid, Khizar Hayat, Baptiste Magnier

STRAIGHTFORWARD ADAPTATION OF PARTICLE FILTER TO FISH EYE IMAGES FOR TOP VIEW PEDESTRIAN TRACKING

Hicham Talaoubrid[1,3], *Khizar Hayat*[2] *and Baptiste Magnier*[3]

[1]L2TI, Université Sorbonne Paris Nord, Paris, France
[2]University of Nizwa, Sultanate of Oman
[3]EuroMov Digital Health in Motion, Univ Montpellier, IMT Mines Ales, Alès, France

Hicham.Talaoubrid@mines-ales.org, khizar.hayat@unizwa.edu.om, Baptiste.Magnier@mines-ales.fr

ABSTRACT

Fisheye lenses are renowned for their capacity to capture incredibly broad perspectives, often reaching up to 180 degrees. Their versatility extends beyond computer vision tasks and encompasses various fields. While applying computer vision techniques directly to fisheye images can yield suboptimal results, this article aims to introduce an uncomplicated and straightforward adaptation of the well-known Bayesian based particle filter. Through a few minor adjustments, we will demonstrate the potential to enhance the particle filter's performance when dealing with such images, particularly for tracking purposes. In this study, we investigate the applicability of particle filters based on features such as: color space, Local Binary Pattern, Histogram of Oriented Gradients, and their combinations. Eventually, experiments and evaluations were carried out on hand annotated real videos with a top view fisheye camera in motion.

Index Terms— Fisheye camera, particle filter, histograms, object tracking

1. INTRODUCTION

Tracking with fisheye cameras presents challenges due to extreme wide-angle views, distortion, and complex perspective transformations, as detailed in [1, 2]. Particle Filters can play a crucial role in addressing these challenges by providing a flexible and adaptable framework for tracking objects in non-linear, wide-angle scenes. Indeed, they can handle the unique characteristics of fisheye images, such as distortion correction and probabilistic state estimation, making them a valuable tool for robust object tracking in fisheye camera setups.

Recent deep-learning-based techniques are efficient to detect and track people, even using fisheye sensors [3, 4, 5, 6, 7, 8, 9] require large datasets, may struggle with non-linearity, and lacks the same level of interpretability. Particle Filters (PF) offer advantages over deep learning approaches in scenarios with limited data, nonlinear systems, and the need for interpretability. Often referred to as Sequential Monte Carlo methods [10], they are a class of probabilistic algorithms used for state estimation and tracking in various fields, including computer vision, robotics, and signal processing [11, 12]. These filters provide a versatile framework for estimating the state of a dynamic system over time based on noisy observations by testing several hypothesized target positions as a function of extracted features. PF provide transparency in state estimation; they are particularly well-suited for scenarios where the underlying system is complex, nonlinear, and exhibits non-Gaussian behavior [13]. Briefly, PF remain valuable for state estimation and tracking tasks, complementing the capabilities of deep learning in various applications.

This work focuses on a single target with a slightly moving camera. Getting a usable dataset is difficult in this very specific field, so we filmed and annotated videos that we will publicly release. The developed approach is based on a particle filter and the features correspond to color space [14], Local Binary Pattern (LBP) [15] and Histogram of Oriented Gradients (HOG) [16]. As the appearance of the pedestrians vary strongly passing in front of the video, we show the reliability and efficiency of the weights given to these features, depending of the position in the image, as a consequence, the HOG differs too, as illustrated in Fig. 1.

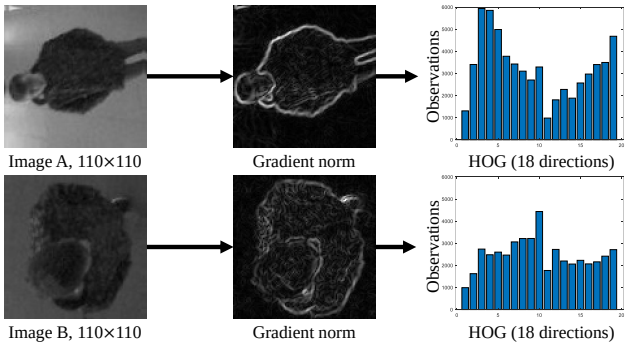
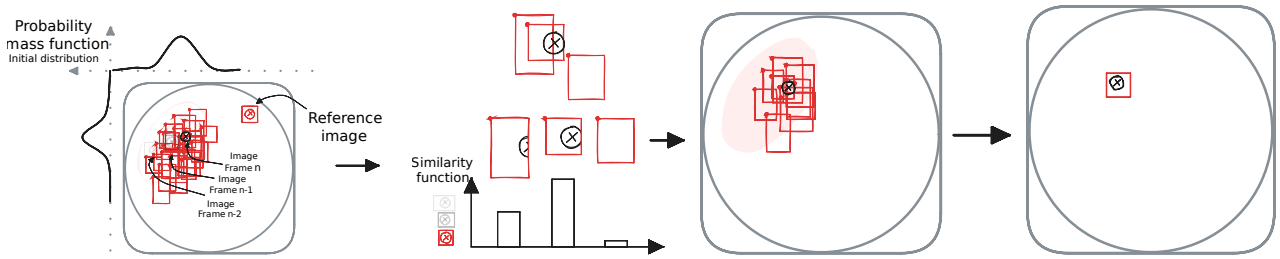


Fig. 1. Example of a pedestrian appearing in a fisheye image in gray level, its gradient norms and tied HOGs. The Image A is on the border of the original image whereas the Image B corresponds to a person passing in front of the focal point.



(a) Sampling (b) Weighting (c) Re-sampling (d) Prediction

Fig. 2. Sequential steps used by PF for tracking. (a) Particles are initialized from prior (position at frame t-1), (b) Weights are updated (detailed in Secs. 3.2.1 and 3.2.2). (c) Particles are resampled based on the calculated weights for a prediction in (d).

2. PARTICLE FILTERING FOR TRACKING

The Particle filter is a widely utilized method for tracking applications. It employs a Bayesian framework where each particle represents the probability of locating the desired object based on particular features, as outlined in [17, 12]. The central concept involves representing the posterior density through a collection of random particles, each assigned specific weights. Subsequently, estimated positions are calculated using these samples and their corresponding weights.

Particle filtering, also known as the Sequential Monte Carlo method [18], has been widely documented in the literature [10, 19] and is summarized in Fig. 2. The stability of the filter hinges on the number of particles. On the one hand, increasing the number of particles enhances filter stability [13]. On the other hand, it leads to longer compilation times. The object tracking implementation here relies on a combination of color histograms [17] and LBP texture features [15] using a combination of the Bhattacharyya coefficient [20] and Simpson index [21], as detailed in the following section.

3. FISHEYE PARTICLE FILTER FOR TOP VIEW PEDESTRIAN TRACKING

When working with top-view fish-eye images, the particle filter must be adapted for tracking; different characteristics can be used, and clever ways of combining them.

3.1. Extracted Features and their Combinations

The developed particle filters exploit various features for object tracking, including color histograms, LBP or HOG, and then explore combinations of these features.

3.1.1. Color Spaces

The choice of color space definitely plays a role in the performance of the particle filter, significantly influencing its ability to accurately track objects in complex visual [11]. Color spaces are representations of color information. Selecting an appropriate color space can greatly enhance object tracking in various scenarios. Different color spaces [22] - such as RGB (red, green, blue), HSV (hue, saturation, value) or CIE L*a*b* (denoted Lab) - offer unique advantages and features that address specific tracking challenges. The choice of the color space affects how the particle filter interprets and matches pixel values, impacting the tracker's ability to distinguish objects from their backgrounds. As conducted in [14], Lab is the more powerful color space for people matching in fisheye images.

3.1.2. Histogram of Oriented Gradient (HOG)

The HOG is a descriptor that relies on the shape and contours of the object [16]. Technically, each image map is partitioned into small fix-sized dyadic cells (8×8 or 16×16 or more) and HOG descriptors would be computed for each cell from its orientation map depending on both the orientation and the magnitude of gradient. The histogram is built from a frequency table based on the orientation angles.

3.1.3. Local Binary Pattern (LBP)

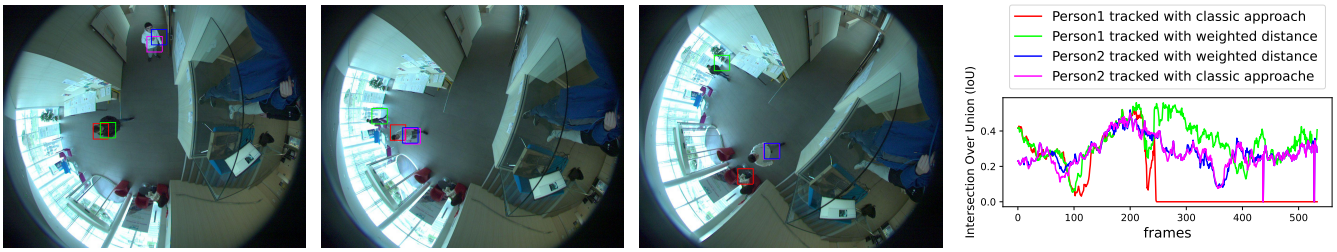
The LBP labels the pixels by comparing each pixel in gray level with its neighborhoods and considering the result as a binary number [15]. Utilizing LBP labels, a histogram can effectively characterize both texture and object shape, making it valuable for tracking purposes [17].

3.1.4. Histogram length

The histogram length depends on the discretization of the color space, the HOG and the LBP. In our tests, the histogram tied to Lab color space contains 24 (*i.e.* 8×3) bins, 18 bins to the HOG and 8 bins appear in the LBP histogram. They can be utilized separately or merged by concatenation according to the proposed method detailed bellow. As an example, LBP and RGB color space histograms are concatenated in [17].

3.2. Adapting to a fish eye camera

Fisheye overhead images are employed, resulting in distortions that pose challenges for tracking when pedestrians move across different image areas. Consequently, continuous tracking becomes unfeasible unless the tracked individual remains within the same image region. When limiting the tracking approach to color histograms, as the tracked person nears the image center, the clothing colors become less discernible,



(a) Frame 56 (b) Frame 251 (c) Frame 300 (d) *IoU* curves over the video

Fig. 3. Detection example with the new weighted distance (in green and blue) compared to the conventional approach.

leaving primarily the top of their head and shoulders visible (refer to Fig. 3). A similar issue arises with the object's gradient, as the target's shape undergoes changes while converging toward the image center, causing the HOG representation to flatten (similar to Image B in Fig. 1).

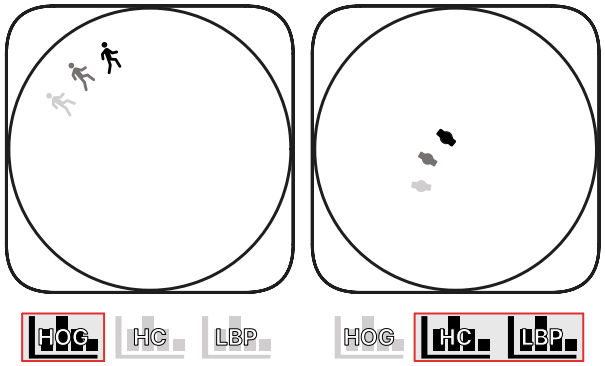
3.2.1. *Changing the target throughout tracking*

We propose to address this issue by defining a new histogram of features as a reference, *i.e.* that will be compared to all the particles' histogram in order to compute the probability density, necessary to perform tracking. This way the particles' features will be compared to the features of updated object that are closer to the actual state of the object.

Thus, to estimate the position of the tracked object in the frame $t+1$, we define the new target histogram as follows:

$$H_{t+1} = ((1-\alpha-\beta)\cdot H_{ref}) + (\beta \cdot H_t) + (\alpha \cdot H_{t-1}), \quad (1)$$

where H_{t+1} is the target histogram for the frame $t+1$, H_t and H_{t-1} are respectively the histograms for frames t and $t-1$. Also, H_{ref} is the histogram of the tracked object at frame 0 and given as an input for the tracking. Finally α and β are two parameters defined empirically (in our tests, α=0.3 and β=0.2). A pipeline optimizing these parameters could be interesting in any future works. To our knowledge, no reference exists about this subject.



(a) Person in the image edge (b) Person in the image center

Fig. 4. Depending if the tracked person is in the center or in the edge of the fish-eye image particle similarity function will base more on HOG or in HC and LBP using a weighted sum. Weights are calculated using Simpson index.

3.2.2. *Weighted contribution from HOG: Simpson Index*

Despite the above function, the tracker still loses the target near the center. By assuming that with fisheye top view cameras, when pedestrians pass close to the center so that only the head top and shoulders are visible, the HOG gives much less information than it should (the signal is relatively flat).

In this way, the HOG makes more sense and is more useful when the tracked objects stay on the image perimeter, while the most meaningful descriptor to use near the center is the color histogram. Consequently, we propose to weight the distances between histograms with the distance of the particles from the center of the image. In that respect, the closer the particles are to the center, the more weight both the color and the texture histogram will carry, and the fewer particles will be compared using the HOG (as represented by the flowchart presented in Fig. 4). They weights are calculated as a function of the Simpson index [21] (see Eq. (2)).

There exists a certain number of functions to discriminate between highly specific (signals containing thin and high peaks) or sparse histograms (flat signals). In this context, the Simpson index is more reliable, as compared in [23]. Let $\{h_k\}_{k=1\ldots N}$ be the set of values of a histogram H built on a partition made of N bins with n observations (*i.e.* $n = \sum_{k=1}^{N} h_k$). The Simpson index $\mathscr{S}$ is defined by:

$$\mathscr{S}(H) = \frac{\sum_{k=1}^{N} h_k \cdot (h_k - 1)}{n \cdot (n-1)}. \quad (2)$$

The most peaked histogram will lead to maximizing $\mathscr{S}(H)$ with large values [21]. Consequently, as the HOG contains more peaked histograms on the image border, the weights are computed as a function of the index $\mathscr{S}$ and the particle coordinates. For this to happen, a function tending to 1 when the HOG signal is flat and to 0 when the HOG contains sharp bins could be: $\mu = e^{1-\mathscr{S}(\text{HOG})}$.

Now, let's assume that H_F represents the feature histograms taken as references (H_F is the concatenation of the color histogram in a specific color space and the LBP labels). Let $p^{(i)}$ be the i^{th} particle, and let $H_{OG}^{(i)}$ and $H_F^{(i)}$ be respectively histogram of oriented gradient and feature histogram of the particle. In order to compare particles' histograms to the target ones, we propose the following distance:

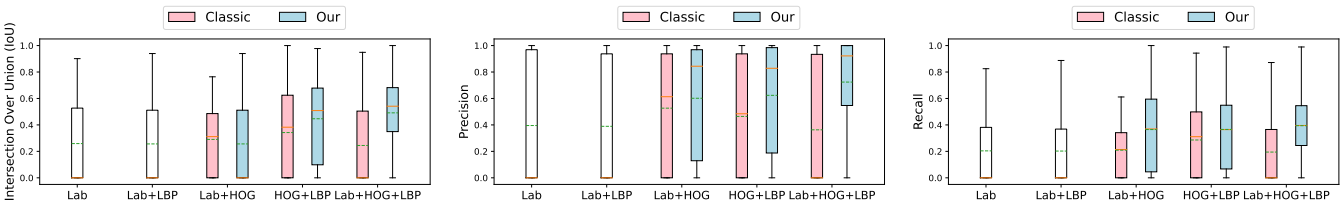


Fig. 5. Comparison of Box Plot presenting *IoU*, *Prec* and *Rec* computed on all the videos weighted by the number of frames.

$$d = \mu \cdot \mathscr{B}\left(H_{OG}^{\text{ref}}, H_{OG}^{(i)}\right) + (1-\mu) \cdot \mathscr{B}\left(H_{F}^{\text{ref}}, H_{F}^{(i)}\right), \quad (3)$$

with $\mu = e^{1-\mathscr{S}(\text{HOG})}$, representing the weight contribution wanted for the HOG histograms. Here, $\mathscr{B}$ represents the Bhattacharyya distance [20] defined by:

$$\mathscr{B}(H_1, H_2) = \sqrt{1 - \frac{1}{\sqrt{\bar{H}_1 \cdot \bar{H}_2 \cdot N^2}} \sum_{k=1}^{N} H_1(k) \cdot H_2(k)}, \quad (4)$$

with N the number of bins in the normalized histogram $H_{i,i\in\{1,2\}}$ and $\bar{H}_{i,i\in\{1,2\}} = \frac{1}{N}\sum_{k=1}^{N} H_i(k)$.

4. EXPERIMENTAL RESULTS AND EVALUATIONS

4.1. Experimental protocol: moving fisheye camera

The proposed tracking method is evaluated and compared with the performance of the basic particle filter (having HOG, LBP or/and color space(s) as features). To do this, we used our own dataset containing 6 videos filmed with a moving top view fisheye camera. The videos contain a total of $5,023$ frames which are annotated. In the presented tests, 50 particles are distributed according to a normal distribution around the tracked object with a standard deviation of 5.

Several well-known metrics are computed to report the reliability of the methods, namely Intersection over Union (*IoU*), Precision (*Prec*) and Recall (*Rec*), which represent confusion matrix-based evaluations, pixel per pixel by comparing the ground truth and the detected object.

4.2. Evaluation and Comparison

Fig. 5 displays box plots representing the three measures, computed across all videos and weighted by the frame count. These plots employ color coding, with the orange line indicating the median value and the green line representing the mean. The performance evaluation of the particle filter is presented for different histogram combinations: Lab, LBP, and HOG. Notably, the incorporation of HOG into the particle filter algorithm significantly enhances its overall performance.

For the basic particle filter, which relies solely on the Lab color space and even when augmented with the LBP histogram, the median performance score is observed to be 0. This outcome suggests that the tracker frequently loses the target object earlier compared to other histogram combinations. However, the introduction of HOG demonstrates superior tracking capabilities, thanks to our novel histogram comparison methodology, particularly when the tracked object does not pass near the center of the image. In such scenarios, the basic approach suffices for satisfactory tracking. In contrast, the tracker using the basic approach loses track of the target in other instances, while the proposed approach continues to perform effectively, as illustrated in Fig. 3. Consequently, the weighted distance metric outperforms the conventional method in these scenarios, leading to a significant improvement in overall performance. This improvement is reflected in an almost 50% increase in the average Intersection over Union (*IoU*) in the video set. It is also worth noting that with the proposed method, the tracked object is only rarely lost, which is remarkable given the challenges posed by distortions, changes in object size, and camera movements. When the tracker does lose the object, it rapidly recovers it.

Finally, it's worth mentioning that the video results are available at the address provided in [24].

5. CONCLUSION AND FUTURE WORKS

Tracking with a top-view fisheye camera presents considerable complexity, particularly when applying deep learning techniques, due to the scarcity of available annotated data. The particle filter remains an interesting alternative; however, the conventional version may fall short in the specific context of our scenario. In this article, we delve into the multitude of features that can be employed with the particle filter, suggesting innovative combinations. The proposes technique is evaluated via three metrics (*IoU*, *Prec* and *Rec*) on videos containing a total of $5,023$ frames. The performance results clearly demonstrate the superiority of our approach when compared to conventional methods, as exemplified in [17], particularly within the context of top-view fisheye images. The proposed distance is employed to compare particles' histograms, using the shape of the HOG. Depending on the position of the tracked object, the elements of its body will be more or less apparent, and so its HOG will be more or less flat. By computing the Simpson index, the flatness of this histogram is evaluated, by assigning various weights to the HOG coupled with the other features. Indeed, it effectively addresses the distortion challenges encountered when the tracked object comes into proximity to the image center.

To conclude, based on the works in [25] and [19], it is possible to run the particle filter in real time with limited computing resources. Furthermore, the proposed Simpson-index technique can be combined with a neural network to enhance specific object detection/tracking tasks.

6. REFERENCES

[1] J. Kumler and M. Bauer, "Fish-eye lens designs and their relative performance," in *Current developments in lens design and optical systems engineering*. SPIE, 2000, vol. 4093, pp. 360–369.

[2] S. Gao, K. Yang, H. Shi, K. Wang, and J. Bai, "Review on panoramic imaging and its applications in scene understanding," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–34, 2022.

[3] J. Yu, A. C. Perez Grassi, and G. Hirtz, "Applications of deep learning for top-view omnidirectional imaging: A survey," *arXiv preprint arXiv:2304.08193*, 2023.

[4] D. Fuertes, C.R del Blanco, P. Carballeira, F. Jaureguizar, and N. García, "People detection with omnidirectional cameras using a spatial grid of deep learning foveatic classifiers," *DSP*, vol. 62, pp. 291–294, 2022.

[5] Q. N. Minh, B. Le Van, C. Nguyen, A. Le, and V.D. Nguyen, "ARPD: Anchor-free rotation-aware people detection using topview fisheye camera," in *IEEE AVSS*, 2021, pp. 1–8.

[6] Z. Duan, O. Tezcan, H. Nakamura, P. Ishwar, and J. Konrad, "Rapid: Rotation-aware people detection in overhead fisheye images," in *IEEE/CVF CVPR Workshops*, 2020, pp. 636–637.

[7] Y.-C. Huang, C.-W. Liu, and J.-H. Chuang, "Using fisheye camera for cost-effective multi-view people localization," in *IEEE ICIP*, 2021, pp. 3248–3252.

[8] O. Haggui, H. Bayd, and B. Magnier, "Centroid human tracking via oriented detection in overhead fisheye sequences," *The Visual Computer*, pp. 1–19, 2023.

[9] N. Odic, B. Faure, and B. Magnier, "FORT: Fisheye Online Realtime Tracking," in *IEEE MMSP*, 2023, pp. 1–6.

[10] M. Isard and A. Blake, "Condensation—conditional density propagation for visual tracking," *IJCV*, vol. 29, no. 1, pp. 5–28, 1998.

[11] C. Teulière, E. Marchand, and L. Eck, "A combination of particle filtering and deterministic approaches for multiple kernel tracking," in *IEEE ICRA*, 2009, pp. 3948–3954.

[12] A.A. Mekonnen, F. Lerasle, and A. Herbulot, "Cooperative passers-by tracking with a mobile robot and external cameras," *CVIU*, vol. 117, no. 10, pp. 1229–1244, 2013.

[13] P. Closas and C. Fernández-Prades, "Particle filtering with adaptive number of particles," in *IEEE Aerospace Conference*, 2011, pp. 1–7.

[14] H. Talaoubrid, M. Vert, K. Hayat, and B. Magnier, "Human tracking in top-view fisheye images: Analysis of familiar similarity measures via HOG and against various color spaces," *J. of Imaging*, vol. 8, no. 4, pp. 115, 2022.

[15] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE TPAMI*, vol. 24, no. 7, pp. 971–987, 2002.

[16] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE CVPR*, 2005, vol. 1, pp. 886–893.

[17] M.D Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. Van Gool, "Robust tracking-by-detection using a detector confidence particle filter," in *ICCV*. IEEE, 2009, pp. 1515–1522.

[18] J.S Liu and R. Chen, "Sequential monte carlo methods for dynamic systems," *J. of the American Stat. Assoc.*, vol. 93, no. 443, pp. 1032–1044, 1998.

[19] O. Haggui, M. A. Tchalim, and B. Magnier, "A comparison of OpenCV algorithms for human tracking with a moving perspective camera," in *IEEE EUVIP*, 2021, pp. 1–6.

[20] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE TPAMI*, vol. 25, no. 5, pp. 564–577, 2003.

[21] E.H Simpson, "Measurement of diversity," *Nature*, vol. 163, no. 4148, pp. 688–688, 1949.

[22] N.A Ibraheem, M.M Hasan, R.Z Khan, and P.K Mishra, "Understanding color models: a review," *ARPN Journal of science and technology*, vol. 2, no. 3, pp. 265–275, 2012.

[23] B. Magnier, F. Comby, O. Strauss, J. Triboulet, and C. Demonceaux, "Highly specific pose estimation with a catadioptric omnidirectional camera," in *IEEE IST*, 2010, pp. 229–233.

[24] H. Talaoubrid and B. Magnier, "Particle filter result," https://partage.imt.fr/index.php/s/Kc8f5esabHXbFkP, 2024.

[25] M. Ullah, F.A. Cheikh, and A.S. Imran, "HoG based real-time multi-target tracking in Bayesian framework," in *IEEE AVSS*, 2016, pp. 416–422.