

Leader and breakaway detection in racing sports videos

Sotirios Papadopoulos
Dept. of Informatics
Aristotle University of Thessaloniki
Thessaloniki, Greece
papadops@csd.auth.gr

Charalampos Symeonidis
Dept. of Informatics
Aristotle University of Thessaloniki
Thessaloniki, Greece
charsyme@csd.auth.gr

Ioannis Pitas
Dept. of Informatics
Aristotle University of Thessaloniki
Thessaloniki, Greece
pitas@csd.auth.gr

Abstract—This paper addresses the important problem of leader detection in racing sports videos (e.g., cycling, boating and car racing events), as his/her proper framing is a pivotal issue in racing sports cinematography, where the events have a linear spatial deployment. Over the last few years, as autonomous drone vision and cinematography emerged, new challenges appeared in drone vision. While, until recently, most computer vision methods typically addressed still camera AV footage, drone sports cinematography typically employs moving cameras. In this paper, we solve the problem of leader detection in a group of similarly moving targets in sports videos, e.g. the leader of a sports cyclist group and his/her breakaway during a cycling event. This is very useful in drone sports cinematography, as it is important that the drone camera automatically centers on such a leader. We demonstrate that the novel method described in this paper can effectively solve the problem of leader detection in sports videos.

Index Terms—UAV cinematography, computer vision, target tracking, event detection

I. INTRODUCTION

As drone sports cinematography evolves over time, there is a constant need for automation. Sports filming tasks, which previously required a human operator, have reached the stage where quick decisions are compulsory and automation seems to be inevitable. When filming a group of moving targets (e.g., athletes) moving in the same direction, in conventional human-operated cinematography, the operator usually targets the camera either on the gravitational center of the athlete group or on the leader of this group. In this sense, leader target detection would enable automatic leader framing by the drone cinematography camera.

The novel leader detection method proposed in this paper, uses global optical flow, in order to estimate camera motion direction. The underlying assumption is that the drone already follows the athlete group, either from above or from a lateral position, according to the chosen drone cinematography mode [1]. A visual target (object) detector and tracker is employed for finding regions of interest (ROI) of the targets (athletes) on the image plane. In the next step, the target ROI centers are projected on the optical flow unit direction vector and, lastly, the leading target (athlete) is detected. Moreover, the

athlete winning order (1st, 2nd, 3rd, etc), as well as their spatial distribution over time can be determined as well, thus providing very useful information for computational racing sports coaching.

Furthermore, in racing sports, a breakaway is the event where, starting from a spatially compact racer group, one athlete accelerates and distances fast from the rest of the athlete group. In the context of this paper, we solve the breakaway detection problem as well, by introducing additional target breakaway detection metrics and constraints.

II. RELATED WORK

Although there have been numerous works that study athletic performance at an individual athlete level, primarily by modeling biomechanics [2]–[4], there has been no research on the kinematics and dynamics of racing sports involving a group of athletes using only visual information, at least to the authors' knowledge.

III. LEADER DETECTION

The first step to leading athlete (leader) detection is visual athlete detection on a video frame, producing athlete ROIs, as shown in Figure 1. Alternatively, semantic segmentation [5]–[7] can be utilized for better localization accuracy. Then global motion is estimated, which is primarily due to drone and ensuing drone cinematography camera motion.

In such a framework, target group motion direction estimation can be obtained by assuming that the drone camera moves on a 3D line that is parallel to the 3D target trajectory, while maintaining the camera axis to be almost perpendicular to the 3D drone motion direction, trying to keep the group of racers inside the video frame, as shown in Figure 2. This means that the targets are almost static on the image plane, but there is global 2D motion to the opposite direction. Thus, detecting the global motion direction can help in determining the actual target group motion direction. The following drone cinematography camera motion types can meet these requirements: Lateral Tracking Shot (LTS) and Vertical Tracking Shot (VTS), according to the UAV shot type taxonomy recently defined and formalized in [1], [8] and [9]. This is not a major limitation, as LTS and VTS are the primary drone cinematography motion types used for when the director wants

to frame the leader. The gimbal points at the center of gravity of the athlete group. The drone camera focal length should be fixed. Furthermore, the video framing should be of long shot type [1], so that many targets appear in the video frame.

Video frames can be greatly subsampled, resulting in, e.g., 144×256 pixel images, as this resolution is enough for our goals. Optical flow estimation algorithm [10] has proven to be reliable and fast, as it takes approximately 7.4 ms for optical flow estimation on such subsampled video frames. The 2D target group motion unit direction vector $-\hat{v}$ estimate, shown in Figure 1, is essentially the reverse of the 2D global motion unit direction vector \hat{v} estimate, which can be found by, e.g., averaging the estimated optical flow vectors $\bar{\mathbf{v}} = \frac{1}{N} \sum_{i=1}^N \mathbf{v}_i$. To avoid excessive computational cost, only few optical flow vectors at chosen locations are used for the mean optical flow vector calculation. In order to avoid letting outliers affect \hat{v} , zero motion vectors at target locations are removed, as targets appear static on the image plane. To further eliminate possible outliers, the Local Outlier Factor algorithm [11] is used to detect and remove other possible optical flow outliers. Another possibility is to use either a vector median [12], or angular statistics [13] to find \hat{v} .

The ordering of the targets/athletes and leader detection can be easily achieved by projecting the target ROI centers on $-\hat{v}$, as depicted in Figure 1.

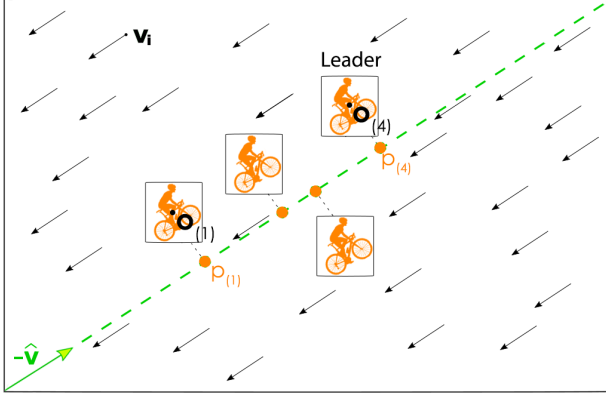


Fig. 1: Projection of the targets on the estimated global motion direction vector $-\hat{v}$.

IV. LEADER BREAKAWAY DETECTION

Let $\mathcal{C} = \{c_1, \dots, c_n\}$ denote a group of n targets. Their projections on $-\hat{v}$ are denoted by $p_i = -\mathbf{o}_i^T \cdot \bar{\mathbf{v}}^T$, $i = 1, \dots, n$, where $\mathbf{o}_1, \dots, \mathbf{o}_n$ denote the corresponding target ROI center coordinates on the image plane. Let $c_{(1)} \preceq c_{(2)} \preceq \dots \preceq c_{(n)}$ be their ordered rank [12] along the motion direction, $\mathbf{o}_{(1)}, \mathbf{o}_{(2)}, \dots, \mathbf{o}_{(n)}$ their corresponding coordinates and $p_{(1)}, p_{(2)}, \dots, p_{(n)}$ their corresponding projections on $-\hat{v}$. $c_{(n)}, c_{(1)}$ denote the leader and the laggard, respectively. During breakaway, the distances of the leader location $\mathbf{o}_{(n)}$ from either the mean target group location $\bar{\mathbf{m}} = \frac{1}{n} \sum \mathbf{o}_i$

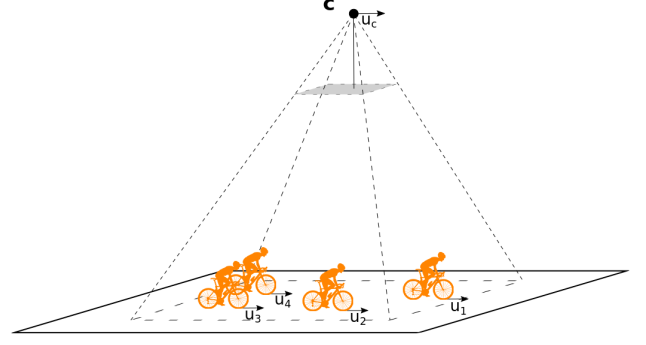


Fig. 2: Proper UAV cinematography camera motion with respect to the motion of the targets.

and/or from the second target $\mathbf{o}_{(n-1)}$ increases sharply over time, thus indicating breakaway. Therefore, we propose the following metrics that can indicate the breakaway occurrence.

Distance to variance ratio has two variations. The first one is the ratio of the squared distance between the first target location $\mathbf{o}_{(n)}$ and the mean $\bar{\mathbf{m}}$ of target ensemble $\mathcal{C} - \{c_{(n)}\}$ coordinates divided by its variance:

$$r_1 = \frac{\|\mathbf{o}_{(n)} - \bar{\mathbf{m}}\|_2^2}{\sum_{\mathbf{o} \in \tilde{\mathcal{O}}} \|\mathbf{o} - \bar{\mathbf{m}}\|_2^2}, \quad (1)$$

where $\tilde{\mathcal{O}}$ contains the target coordinates corresponding to $\mathcal{C} - \{c_{(n)}\}$.

The second one is the ratio of the squared distance between the leader target location $\mathbf{o}_{(n)}$ and the second leader location $\mathbf{o}_{(n-1)}$ divided by the variance of the target ensemble $\tilde{\mathcal{O}}$:

$$r_2 = \frac{\|\mathbf{o}_{(n)} - \mathbf{o}_{(n-1)}\|_2^2}{\sum_{\mathbf{o} \in \tilde{\mathcal{O}}} \|\mathbf{o} - \bar{\mathbf{m}}\|_2^2}. \quad (2)$$

However, these metrics have a limitation. They assume that the spatial target location projection variance along the motion direction $\sum_{\mathbf{o} \in \tilde{\mathcal{O}}} \|\mathbf{o} - \bar{\mathbf{m}}\|_2^2$ is almost static. This is not always true, since targets can move independently, hence altering r_1 and r_2 drastically, without an actual breakaway event occurrence. Additionally, laggards affect target group position variance, hence r_1, r_2 as well. Finally, both r_1, r_2 depend on $\bar{\mathbf{m}}$, whose estimation can become unreliable if targets start disappearing from the video frame, due to a large target group spatial spread or due to changes in camera focal length.

Distance to distance ratio was created as a solution to the target disappearance problem, as well as to counter the effects of the laggards, as both the leader and the laggard positions are essentially outliers of the target position probability density functions (pdfs). The first variation is the ratio of the squared



Fig. 3: Cyclist ranking. The first row shows the source video frames and the second row shows the processed ones. The orange line indicates the estimated target motion direction. The yellow, magenta and blue dots stand for a) the leader, b) the second best and c) the rest of the ranked targets respectively.

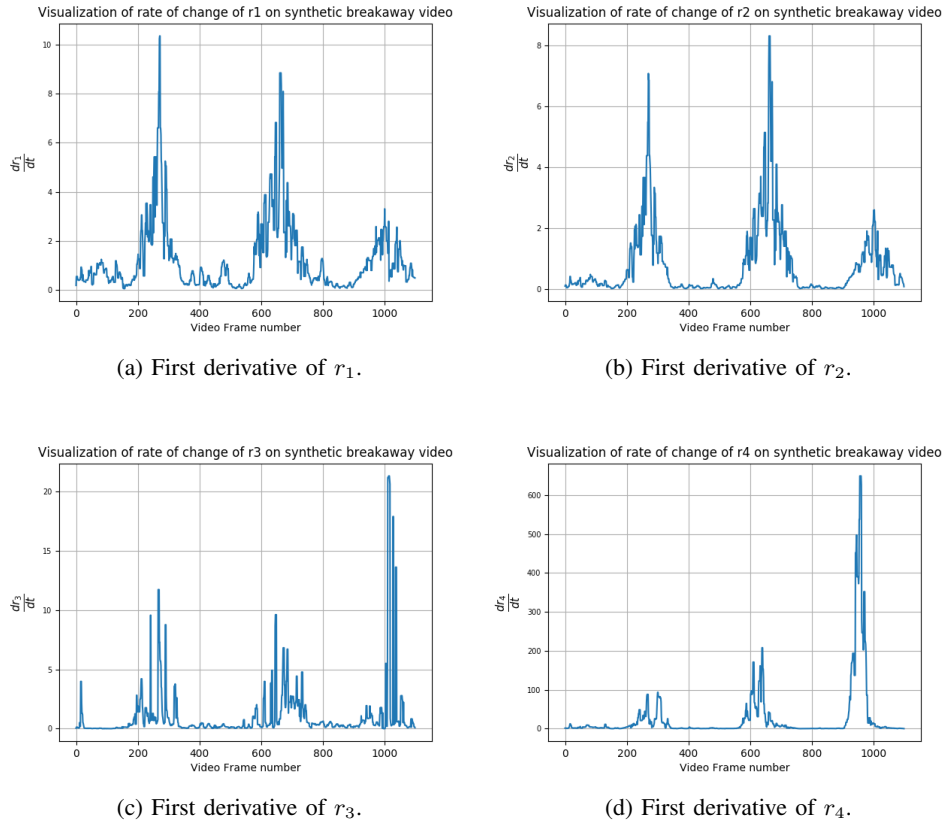


Fig. 4: These figures present the first derivatives of r_1, r_2, r_3 and r_4 through time on a simulated video.

distance between the leader and the second leader target, divided by median absolute deviation (MAD) in $\tilde{\mathcal{O}}$ [12]:

$$r_3 = \frac{\|\mathbf{o}_{(n)} - \mathbf{o}_{(n-1)}\|_2^2}{\text{median}(\|\mathbf{o}_{(i)} - \text{med}_2(\mathbf{o}_{(i)})\|_2)}, \quad i = 1, \dots, n-1, \quad (3)$$

where $\text{med}_2()$ can be any 2D median operator, e.g., a marginal or a vector median [12] [14].

The second variation is the ratio of the squared distance $\|\mathbf{o}_{(n)} - \mathbf{o}_{(n-1)}\|_2^2$, divided by mean distance between each target and its nearest neighbor:

$$r_4 = \frac{\|\mathbf{o}_{(n)} - \mathbf{o}_{(n-1)}\|_2^2}{\frac{1}{|\tilde{\mathcal{O}}|} \sum_{\mathbf{o} \in \tilde{\mathcal{O}}} \|\mathbf{o} - \underset{\mathbf{k} \in \tilde{\mathcal{O}}}{\text{argmin}}(\|\mathbf{o} - \mathbf{k}\|_2^2)\|_2^2}. \quad (4)$$

Both r_3, r_4 assume that target ROI center variance along the

direction perpendicular to \hat{v} is small. If not, the same analysis can be performed on $p_{(i)}$. In this case, the med_2 operator in (3) is reduced to the classic 1D median operator [12].

All r_i , $i = 1, \dots, 4$ vary with time t . As leader breakaway is a temporal event, their differentiation indicates breakaway: $\frac{dr_i}{dt} > T$, $i = 1, \dots, 4$. The optimal threshold T can be found by estimating the pdfs $P_{r_i}(r_i)$, $i = 1, \dots, 4$ in presence/absence of breakaway and by choosing a threshold minimizing the detection error [15].

TABLE I: Standard deviation of breakaway metrics in absence of breakaway

Metric	Std in absence of breakaway
r_1	0.18
r_2	0.17
r_3	0.06
r_4	0.14

V. EXPERIMENTAL RESULTS

Leader and breakaway performance was evaluated in a video dataset of cycling races provided by Radiotelevisione Italiana (RAI) ¹. From the whole dataset, a number of relatively short video clips was extracted, so that the drone camera motion conforms to the detection requirements set up in the previous sections. The cyclist detector [16] was employed for target detection. The chosen video clips were transformed to grayscale and their spatial resolution was reduced to 144×256 pixels.

Due to the fact that the leader detection algorithm uses bounding boxes that have been already produced by a cyclist detector, the quality of the overall results depends strongly on the target detection performance. If target detection performance is good [16], as in 3, the leader detection algorithm yields almost perfectly accurate results. Using a dataset of 1571 bicycle racing video frame pairs, the leader detection algorithm achieved a high leader detection accuracy of 97.2%, while in detecting the second best it achieved 95.6% accuracy. It can run in real time, as only around 24 ms are needed to process a single video frame pair.

As cyclist breakaway videos are rather scarce, at least to authors' knowledge, the event was simulated. Figures 5 (a) and 5 (b) are taken from a CGI video simulating a cyclist racing sports event shot with VTS CMT [1] that includes three breakaway events. Figure 4 depicts the derivatives $\frac{dr_i}{dt}$, $i = 1, \dots, 4$, respectively, filtered with median filtering of kernel size 7. As seen in this figure, if the derivative surpasses a certain threshold, it can successfully indicate the time instances (around video frames no. 250, 600 and 1000) when the breakaways take place. It is clearly seen that $\frac{dr_3}{dt}$ and particularly $\frac{dr_4}{dt}$ have much less background noise in the absence of a breakaway, as manifested in Table I. As $\frac{dr_3}{dt}$ and $\frac{dr_4}{dt}$ are more robust, the choice of the optimal thresholds for $\frac{dr_3}{dt}$ and $\frac{dr_4}{dt}$ is investigated. First we estimate

$p_{r_3}(r_3)$ and $p_{r_4}(r_4)$ in presence/absence of breakaway using Kernel Density Estimation [17] with Gaussian Kernel and then determine the optimal thresholds that minimize the detection error, as seen in Figure 6.

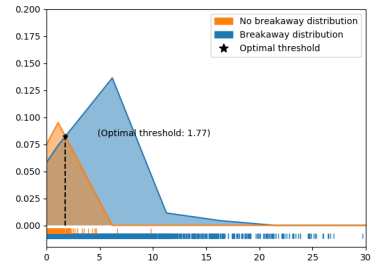


(a) Video frame at a time instance when the cyclist group is compact.

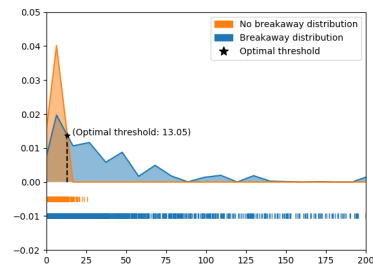


(b) Video frame at a time instance when a breakaway is detected.

Fig. 5: Video frames from a cyclist race simulation video



(a) r_3 .



(b) r_4 .

Fig. 6: Optimal threshold identifications for r_3 and r_4 .

VI. CONCLUSIONS

Although the results show that the leader and breakaway detection algorithm can yield satisfactory results, there are

¹http://www.aiaa.csd.auth.gr/LAB_PROJECTS/MULTIDRONE/AUTH_MULTIDRONE_Dataset.html

some certain limitations that must taken into consideration. First and foremost, the leader detection algorithm applies only in scenarios where the targets move on a relatively straight line. Turns produce complex optical flow behavior, such that the motion direction can not be estimated by a single averaging. Secondly, as stated in a previous section, the algorithm works well only in certain CMTs, notably VTS and LTS [1], [8], [9], where the camera axis is perpendicular to the targets trajectory and the UAV and target trajectories are parallel to one another.

[17] Murray Rosenblatt, "Remarks on some nonparametric estimates of a density function," *The Annals of Mathematical Statistics*, pp. 832–837, 1956.

REFERENCES

- [1] Ioannis Mademlis, Nikos Nikolaidis, Anastasios Tefas, Ioannis Pitas, Tilman Wagner, and Alberto Messina, "Autonomous uav cinematography: A tutorial and a formalized shot-type taxonomy," *ACM Computing Surveys (CSUR)*, vol. 52, no. 5, pp. 105, 2019.
- [2] Ami Drory, "Computer vision and machine learning for biomechanics applications: Human detection, pose and shape estimation and tracking in unconstrained environment from uncalibrated images, videos and depth.," 2017.
- [3] Koen APM Lemmink, Stuart Morgan, Jaime Sampaio, and Dietmar Saupe, "Computer science in high performance sport-applications and implications for professional coaching (dagstuhl seminar 13272)," in *Dagstuhl Reports*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2013, vol. 3.
- [4] Steffi L Colyer, Murray Evans, Darren P Cosker, and Aki IT Salo, "A review of the evolution of vision-based motion analysis and the integration of advanced computer vision methods towards developing a markerless system," *Sports medicine-open*, vol. 4, pp. 24, 2018.
- [5] Sotirios Papadopoulos, Ioannis Mademlis, and Ioannis Pitas, "Neural vision-based semantic 3d world modeling," in *2021 IEEE Winter Conference on Applications of Computer Vision Workshops (WACVW)*. IEEE, 2021, pp. 181–190.
- [6] Sotirios Papadopoulos, Ioannis Mademlis, and Ioannis Pitas, "Semantic image segmentation guided by scene geometry," in *2021 IEEE International Conference on Autonomous Systems (ICAS)*, in press. IEEE, 2021.
- [7] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.
- [8] Ioannis Mademlis, Vasileios Mygdalis, Nikos Nikolaidis, Maurizio Montagnuolo, Fulvio Negro, Alberto Messina, and Ioannis Pitas, "High-level multiple-uav cinematography tools for covering outdoor events," *IEEE Transactions on Broadcasting*, 2019.
- [9] Iason Karakostas, Ioannis Mademlis, Nikos Nikolaidis, and Ioannis Pitas, "Shot type constraints in uav cinematography for autonomous target tracking," *Information Sciences*, vol. 506, pp. 273–294, 2020.
- [10] Gunnar Farnebäck, "Two-frame motion estimation based on polynomial expansion," in *Scandinavian conference on Image analysis*. Springer, 2003, pp. 363–370.
- [11] Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander, "Lof: identifying density-based local outliers," in *ACM sigmod record*. ACM, 2000, vol. 29, pp. 93–104.
- [12] I Pitas and A Venetsanopoulos, "Nonlinear digital filters: Principles and applications, klumer academic," *Publishers Hingham*, 1990.
- [13] Nikolaos Nikolaidis and Ioannis Pitas, "Nonlinear processing and analysis of angular signals," *IEEE Transactions on Signal Processing*, vol. 46, no. 12, pp. 3181–3194, 1998.
- [14] Zvi Drezner and Horst W Hamacher, *Facility location: applications and theory*, Springer Science & Business Media, 2001.
- [15] Sergios Theodoridis and Konstantinos Koutroumbas, "Pattern recognition," 2008, Academic Press.
- [16] Paraskevi Nousi, Ioannis Mademlis, Iason Karakostas, Anastasios Tefas, and Ioannis Pitas, "Embedded uav real-time visual object detection and tracking," in *Proceedings of the IEEE International Conference on Real-time Computing and Robotics (RCAR)*, 2019.