Tiny People Finder: Long-Range Outdoor HRI by Periodicity Detection

Jake Bruce, Valiallah (Mani) Monajjemi, Jens Wawerla and Richard Vaughan
Autonomy Lab, Simon Fraser University
{jakeb, mmonajje, jwawerla, vaughan}@sfu.ca

***Abstract*— We present a novel method for detecting waving humans at long ranges in outdoor environments, using a consumer video camera on a mobile ground-based robot. The proposed algorithm analyzes the average pixel intensity of motion-containing regions in an image stream, identifying those regions which show a strong periodic signal in the frequency range of human waving gestures. The system achieves robustness to sporadic false positives such as waving trees, flags and walking pedestrians by using a filter to reject non-stationary and transient detections. In real-world experiments we determine the effective detection range of our algorithm and show that a robot equipped with a low-resolution consumer camera is able to approach a single waving human from a starting position up to 35 meters away when the person is roughly 20 pixels high, even in the presence of human and non-human periodic distractors such as foliage.**

I. INTRODUCTION

Consider the following situation: you are standing on a hill looking down into a crowd of people around a hundred meters away, attempting to find a friend. You are too far away to see any faces, and you don't know what color clothing she is wearing today. As you scan the scene, suddenly a repetitive motion catches your eye: you see your friend in the middle of the crowd, waving at you with both arms.

In a cluttered scene like the one described above, even the human visual system can fail to locate target objects until presented with a hint in the form of a color or a salient motion. In scenes that already contain a lot of motion, a repetitive action like a waving gesture can serve as a crucial clue to help direct the attention of the seeker. The person who wants to be found is injecting an unusual salient signal into the visual field of the seeker.

We are interested in methods for robots to cooperate with humans in large-scale outdoor environments. A useful component is to able to identify and approach humans over long distances where people can be as small as 20 pixels high, and against moving and cluttered backgrounds (see Figs. 1, 2). We demonstrate a human-robot interaction (HRI) system that uses consumer camera hardware to detect periodically oscillating image regions and identify candidate humans from long distances, after which the robot approaches the target for close-range interaction. Detections are based on periodic variation in pixel intensity over time, so we need make no assumptions about skin or clothing color, the texture of the background, or the precise scale of the human.

The main contribution of this paper is a long-range periodic gesture detection algorithm that reliably identifies periodic motions using a low-resolution camera in which the human is roughly 20 pixels high. The HRI system in this paper is the first to our knowledge that can locate and approach uninstrumented gesturing humans composed of so few pixels in indoor and outdoor environments, using only monocular camera sensing.

Fig. 1: View of a 20-pixel tall waving human at 35 meters, which was successfully detected and approached to within 3 meters by the system described in this paper.

II. RELATED WORK

Existing vision-based systems for uninstrumented HRI with low-resolution cameras require humans to be located less than ten meters from the robot to ensure they are composed of enough pixels to be identifiable. This is usually due to the use of face detection [1], [2], [3], skin detection [4], or model-based methods that require identification of particular body parts [5]. Action recognition methods [6] have been developed that operate at relatively long ranges (with humans as small as 30 pixels in height) but these assume a cropped figure-centric bounding box, which is difficult to extract when the human targets are very small or in front of a cluttered background. Discrimination between 24 distinct gestures has been accomplished using frame-to-frame difference images [7] but once again the performance of this method degrades when the humans in the image are very small and dominated by noise. Optical flow-based techniques ([8], [9]) that rely on sparse features also tend to be unreliable for gestures at very long ranges, and computing real-time dense optical flow is not feasible on our robot due to limited computational resources.

The detection of periodic signals in image data has been under investigation for more than twenty years in the computer vision community, and a rich collection of approaches have been proposed. Some methods ([10], [11], [12]) track specific points or objects as they move through image space. Image alignment-based methods assume a figure-centric stabilized bounding box, and compute the self-similarity [13] or match points between periods of the oscillatory motion [14]. [15] makes use of aligned bounding boxes to compute a fast Fourier transform (FFT) of the pixels in the image over time, and fits the resulting frequency spectra to periodicity templates to discriminate oscillating pixels. We have experienced artificial periodicity due to small errors in feature tracking and image alignment methods at long ranges, so we are interested in investigating other approaches.

A pedestrian detection algorithm for infrared and color sensors has been proposed [16] that identifies human gaits using a periodicity metric called a *periodogram* [17], which is a quantitative measure of the degree to which a signal is periodic, based on the strength of the signal response at different frequencies. Periodic signal analysis has also been applied to long-range surveillance video [18] to identify walking pedestrians by analyzing the periodograms of blob trajectories, and by looking for an in-phase relationship between blob size and position. Segmentation-based techniques that rely on distinctive blobs become unreliable at long ranges, and currently a robot is more likely to include a visual camera than an infrared sensor due to relative cost.

Offline approaches have been developed to identify multiple periodic motions in video sequences by whole-video frequency and phase spectrum analysis [19], and to detect two-dimensional perspectives of oscillations in three-dimensional space using principles of affine invariance [20]. Affine invariance has the advantage of handling moving cameras, but these are not real-time methods.

Periodic signals in camera streams were exploited on the *Aqua* underwater robot to track and follow the oscillatory kicking motion of human divers at close range [21]. The FFT is performed on a time series of average pixel intensities in regions of the image, and significant peaks in the desired frequency are identified. A similar approach is described in [22] in which a support vector machine is trained to discriminate between gestures on the basis of frequency and phase spectra in an aggressively downsampled image. Our previous work [23] uses image stabilization, feature clustering under the assumption of an approximately planar environment, and FFT to detect stationary waving gestures on-board a UAV in flight.

We propose a real-time monocular vision method based on the *Aqua* robot system [21] combined with the periodicity metric from [12], which scores moving regions proportional to the relative strength of the fundamental frequency and its harmonics compared to the rest of the spectrum. This paper contributes a vision system that improves the maximum range of the state of the art for detecting waving human gestures by analyzing intensity changes for periodicity in very small regions of the image, and uses a flexible clustering algorithm to identify waving gestures in challenging non-planar environments using only low-resolution consumer camera equipment.

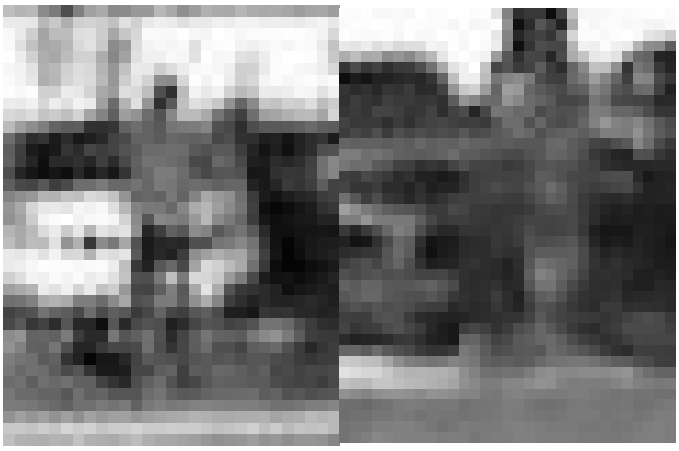


Fig. 2: Two examples of waving humans approximately 20 pixels tall whose gestures were successfully detected.

III. Periodic Motion

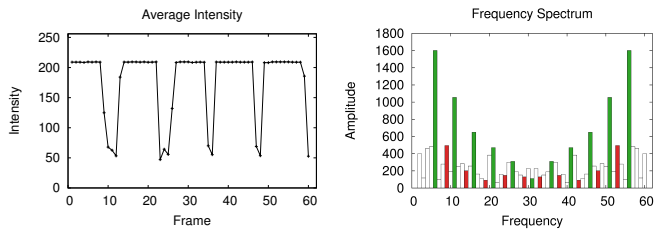
The proposed algorithm for detection of periodic image regions can be described in three stages: *A)* constructing a time series of average pixel intensity on a per-region basis, *B)* identification of periodic signals in the desired frequency range for human waving, and *C)* clustering periodic regions into large-scale bounding boxes for output. This section describes each stage in detail.

A. Intensity Time Series

To construct a set of time series buffers to check for periodicity, we divide the image into a set of regions of interest and compute the average grayscale intensity of the pixels in each region over time. Since we are interested in detecting people on the order of 20 pixels in height (see Fig. 2), we use regions 10 pixels on a side, which we overlap by half along each axis. Using smaller regions increases the range of the detector by allowing it to detect smaller motions, but also increases the computational demand due to greater region count.

Our system requires that the robot be stationary in order for these image regions to remain in place as time goes on. We are currently investigating methods for ego-motion cancellation to detect gestures from flying vehicles [23] that assumes most features are roughly co-planar for the homography calculation, but this assumption does not hold in general for ground vehicles. Assuming the robot is stationary simplifies data association between frames: in future work we will remove this constraint.

A weighted average grayscale intensity of the pixels in each region is computed using a Gaussian kernel centered on the middle of the region. This non-uniform weighting reduces edge effects between regions, and ensures that a pixel that only moves inside one region still produces variation in its



(a) Average pixel intensity signal (b) Frequency spectrum

Fig. 3: Intensity signal extracted from a periodic waving gesture and frequency spectrum with DC component removed. Green frequency components indicate the fundamental frequency and its harmonics (F_{iw}), while red indicates the components halfway between ($F_{iw+w/2}$). The periodicity P_F of this signal is 0.61.

box's average intensity. We use a symmetric Gaussian kernel with $\sigma = 5$.

These weighted averages are stored in a circular buffer with a 2 second time horizon for each region: a typical series is shown in Fig. 3a. The length of the temporal window should be chosen to include at least two periods of the gesture in order for the periodicity to be clearly present in the frequency spectrum. Increasing the length of this window increases robustness to false positives, but slows the response time of the detector.

B. Evaluating Periodicity

A periodic signal is composed of a signal oscillating at a fundamental frequency plus its harmonics. In this application, periodic signals are embedded in noisy time series data, so the system must discriminate time series that contain sufficiently strong periodic signals from those that do not, on the basis of the frequency spectrum of the signal (see Fig. 3b). To make this distinction we use a metric proposed in [12] in which the periodicity P_F of a signal with power spectrum F and highest amplitude frequency w is given by:

$$P_F = \frac{\sum_i F_{iw} - \sum_i F_{iw+w/2}}{\sum_i F_{iw} + \sum_i F_{iw+w/2}} \quad (1)$$

This quantity is a normalized difference between the sum of the power spectrum values at the highest amplitude frequency and harmonics, and the sum of the values at the frequencies halfway between. This yields a score indicating the relative strength of the frequency w and harmonics compared to the rest of the spectrum. Signals with P_F near 1 are highly periodic, and P_F values close to 0 describe signals with little to no regular oscillation.

We consider a signal to represent a potential gesture if $P_F > 0.45$ and w is between 1Hz and 3Hz, as humans tend to wave at approximately this rate. In addition, we only consider signals with fundamental amplitude $F_w > 30$, to avoid false positives due to minor lighting fluctuations and compression artifacts. These criteria define the sensitivity of the system to noisy periodic signals, and the frequency range of the target gesture.

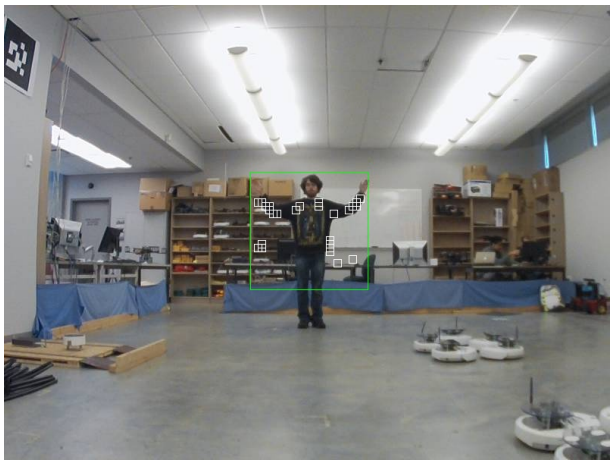


Fig. 4: Periodic regions in white and cluster bounding box in green, produced by the algorithm described in this paper.

We evaluate the periodicity of each time series buffer every N frames, which can be chosen based on available processing power. We choose $N = 15$ on our system for an operating frequency of 2Hz. With more computational power the algorithm can be run at a faster rate, which helps eliminate false positives caused by transient apparent periodicity. We filter the output, requiring a hit-to-miss rate of at least $3:1$ in a region over a 2 second time window before accepting it as a potential stationary waving human. Increasing the length of this window or the required hit-to-miss rate improves robustness to transient periodicity, but slows the response time of the detector.

C. Clustering

Given a collection of 10×10-pixel regions flagged as positive for periodic motion, the system forms large-scale bounding boxes (see Fig. 4) to identify multiple sources of motion if present. We use the scikit-learn implementation [24] of the DBSCAN algorithm [25] which clusters unlabeled data by forming connected subgraphs and makes no assumptions about the number of clusters.

DBSCAN requires two parameters: the maximum distance ϵ between connected data points and the minimum size δ of a cluster. Any detections more than ϵ pixels away from a connected group of at least δ other detections are considered outliers and do not affect the output of the detector. We use $\epsilon = 45$ pixels and $\delta = 3$, chosen to form sensible clusters at both short and long ranges.

IV. Robot Behavior

A detector that only functions while the robot is stationary presents a challenge for defining behavior. We cannot simply servo to the detection, as camera movement can produce apparent periodicity in non-periodic scenes. The position of the stationary detection along the horizontal axis of the image is sufficient to drive our robot accurately in the direction of the target, but how far the robot should travel is not obvious.

The goal of the system is simply to get within range of more discriminative sensors such as face, torso, or human detectors, so the robot drives in the direction of the target for a distance of 10 meters, at which point the robot stops and makes another stationary scan to correct for angle error during detection and approach.

If multiple periodic clusters are detected, we choose a potential human to approach based on whether the gesture persists: walking pedestrians, vehicle traffic, waving flags and trees can all cause apparent periodicity, but intentional waving motion from a human tends to persist while false positives tend to come and go. Our robot behavior waits until exactly one detection is reported before approaching. This also helps at close range, where the two hands of the human can temporarily be clustered as separate periodic motions.

V. Experiments

We evaluate the proposed system using a Husky A200 ground-based robot built by Clearpath Robotics. Excluding emergency-stop behavior, the only active sensor for these experiments is a consumer 640×480 resolution monocular Kinect camera mounted on the front of the robot, providing color video at 30 frames per second. The robot includes an onboard computer with 8GB of RAM and a dual-core Intel Core i5 2.4GHz (2012 laptop-class) processor.

Fig. 5 shows an aerial view of the experimental setting: an outdoor area on Simon Fraser University campus with frequent natural pedestrian and vehicle traffic, waving trees and flags in view (at location F in the image). We perform two sets of experiments: a distance investigation in which detection rates are recorded at different ranges between the robot and four subjects in three locations; and an approach investigation in which the robot attempts to approach to within 3 meters, using the same subjects and locations as the first scenario. Both experiments were repeated against two different backgrounds, shown in Fig. 6, and all trials were performed using the same values for all parameters: the values reported in this paper.

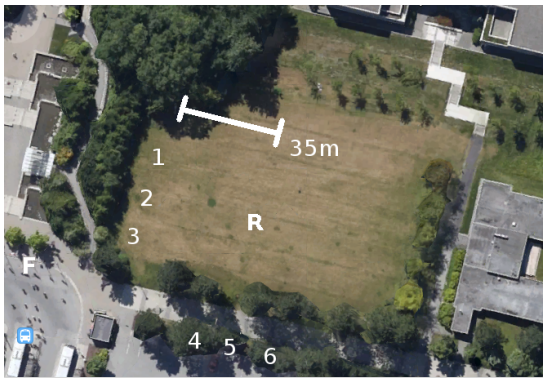


Fig. 5: Outdoor experimental setting on SFU campus. Numbers indicate the locations of waving subjects during experiments, R indicates the initial position of the robot at 35 meters from the subject, and F indicates the location of periodically waving flags.

(a) locations 1, 2, 3

(b) locations 4, 5, 6

Fig. 6: Images from the robot camera with two different backgrounds tested in experiments. The human subjects are barely discernable in the center-left of the images.

Importantly, the human subjects and locations were chosen arbitrarily without testing the detector against these subject/background combinations, to avoid biasing test results. In both distance and approach tests, the human subject performs a two-arm waving gesture as shown in Fig. 4.

A. Distance Experiments

In order to test the effective range of our detector on a 640x480-pixel camera against varying backgrounds, we test whether the system can detect a human waving at distances of 25, 30, 35, 40, and 45 meters. With our camera at these distances, a human is approximately 31, 26, 22, 19, and 17 pixels tall. We evaluate the system at each distance using a stationary camera with four different people standing in three different locations for a total of 12 trials against each of the two background environments. We consider a trial successful as soon as the detector reports a bounding box containing the gesturing human. If the system does not detect the gesture within 60 seconds, the trial is considered a failure. Results are reported in Table I.

	Locations 1,2,3			Locations 4,5,6		
Distance	Success	Time	Rate	Success	Time	Rate
25m	12/12	8.3s	100%	11/12	15.2s	92%
30m	12/12	9.5s	100%	11/12	14.5s	92%
35m	11/12	11.9s	92%	9/12	17.3s	75%
40m	10/12	11.7s	83%	9/12	18.0s	75%
45m	9/12	15.9s	75%	4/12	24.8s	42%

TABLE I: Results of evaluating the maximum range of the detector against two different background environments.

Location	1	2	3	4	5	6
Subject 1	✓	✓	✓	✓	✗	✓
Subject 2	✓	✓	✓	✓	✓	✗
Subject 3	✓	✓	✓	✓	✗	✓
Subject 4	✓	✓	✓	✓	✓	✗
Success rate:	100%			66.6%		
Overall rate:	83.3%					

TABLE II: Results of evaluating the ability of the robot to approach the subject to within 3 meters against different background environments, starting at 35 meters distance.

To analyze the statistical significance of the distance results, we compare against an imaginary detector that chooses a pixel uniformly at random as the center of the detection, and we consider it a successful detection if the chosen pixel is inside the target bounding box. For an SxS-pixel bounding box in a 640x480-pixel image, the probability of success per frame is the ratio of the area of the box to the area of the image.

If we imagine that we match our real trials by running this random detector once every $N = 15$ frames of the video sequences from the distance experiments, the probability that the random detector would successfully identify the waving human over a 60 second window in one of these trials is given by the complement of the probability of not finding the human:

$$p_{trial} = 1 - (1 - \frac{S^2}{640 * 480})^{120} \quad (2)$$

We test the significance of our results against the binomial distribution using $n = 12$ and $p = p_{trial}$ for the size of the human at each distance. We reject with 99% confidence the null hypothesis that our detector is no better than random in every case, except for the 45 meter case with success rate of 4/12, for which we can only say with 85% confidence that our system is better than the imaginary random detector.

B. Approach Experiments

Given an estimate of the success rate of the detector at different ranges, we choose a distance for approach experiments that is likely to find the person but also demonstrate the value of the system in approaching from long range. Our approach experiments are performed with the robot starting 35 meters from the subject, facing toward the middle location in each setting. The robot scans for waving gestures and drives towards the first detection for 10 meters before beginning another scan.

We define an approach as successful once the robot drives to within 3 meters and the person is fully visible in the camera image. The scan-approach behavior continues until the robot stops in front of an obstacle, which will be the human target if the approach succeeds. If the robot does not arrive within 180 seconds, the trial is considered a failure. Results are reported in Table II.

In addition to the waving subject, both environments contained natural and artificial distractors and frequent occlusions of the target. Lighting conditions varied gradually throughout the duration of trials in both environments as blue skies transitioned to clouds. For locations 1, 2 and 3, we planted two stationary humans and four moving humans instructed to wander around the area at varying distances to the robot, often occluding the subject. Also visible in this environment: intermittent bus traffic, waving tree foliage and flags blowing in the wind (2 to 7 kilometers per hour).

For locations 4, 5 and 6, the waving subject was located on the opposite side of a busy pedestrian walkway against a background of parked cars. Most of the distractors in this environment were natural, non-informed human pedestrians who occluded the subject at short intervals, typically between one and five seconds. At times when natural pedestrian traffic was thin, we injected informed humans into the walkway to maintain a roughly consistent occlusion and distraction rate for all trials. In addition to human distractors, this environment contained occasional vehicle traffic and trees moving in the wind, at similar wind speeds as the first set.

See the demonstration video for robot camera views during the experiments: `https://youtu.be/5XmkmdKJ1jY`. The detection task at the larger distances is challenging even for humans.

VI. Discussion

The distance evaluation shows that our system works consistently for smaller pixel sizes than any known real-time method for locating waving humans. It also indicates the effect of contrast to background, and the effect of occlusions and distractors (both natural and human), as the maximum range drops noticeably in locations 4, 5 and 6, and the average time increased considerably over the first environment. This is due both to the challenging background and occlusions of the subject by passing pedestrians, requiring more samples to identify the periodic gesture.

Light and shadow is also a factor, as locations 5 and 6 were partly in shadow which reduced the contrast of the human against the background and resulted in the failure of the approach involving subject 2, who was standing in shadow in front of a dark blue car wearing a blue jacket. This rendered the subject essentially invisible in both grayscale and color images, so the robot did not move at all during this trial.

Shadows can benefit the detector under the right circumstances, as we observed during periods of bright sunlight where the shadow (oscillating along with its human source) exaggerated the size of the apparent periodic motion and

caused it to be detected at greater ranges than would have been possible by the human's appearance alone.

Several failures were due to the conversion of images from color to grayscale. In the approach failures involving subjects 1, 3, and 4, several pedestrians wearing different colored clothing walked by at a roughly constant interval, which can appear periodic in grayscale and caused the robot to drive in the wrong direction. Due to the many-to-one mapping of the grayscale manifold, colors which are visibly distinct in color space are often compressed down to the same or similar grayscale values, as in the case of the red plaid worn by the subject on the right in Fig. 2 against the brown of the tree. This can result in subjects becoming effectively invisible in grayscale even when they are clearly visible in color images, and can make aperiodic color sequences appear periodic in grayscale.

One approach to reducing this effect involves running periodicity checks on all three color channels, detecting regions as positive for periodic motion if any of the channels are moving periodically. Such a system should also reject regions containing color channels with aperiodic but non-stationary signals in any of their color channels. This increased sensitivity can expose the system to false positives which would not have been detected in grayscale however, so we leave this extension for future work.

VII. Conclusions and Future Work

We propose and demonstrate a vision system for long-range HRI: the first system to our knowledge that can locate and approach uninstrumented humans as small as 20 pixels tall, using only a low-resolution consumer camera. Once the robot has approached to close range, traditional interaction techniques become feasible.

Our previous work on long-range HRI investigated video stabilization techniques to allow periodic gestures to be identified from moving cameras [23]. This enables detection during traversal for smoother behavior, and permits the use of this method on aerial vehicles where remaining stationary is rarely an option. The method assumes features are located on a plane approximately parallel to the image plane, but a similar method may be suitable for ground robots with appropriate modification. As mentioned in the discussion section, methods for analyzing periodicity without compressing the color space to grayscale may reduce false negatives caused by the many-to-one mapping to the grayscale manifold.

Other potential improvements include the use of machine learning as shown in [22] to distinguish robustly between human gestures and natural periodicity such as rustling foliage and flags blowing in the wind. Although in practice the approach behavior of our robot mitigates false positives through investigation at close range, discriminating these from afar may help prevent the robot from leaving the area to investigate obvious distractors.

ACKNOWLEDGMENT

This work was supported by the NSERC Canadian Field Robotics Network (NCFRN).

References

[1] K. K. Kim, K.-C. Kwak, and S. Y. Ch, "Gesture analysis for human-robot interaction," in *Advanced Communication Technology, 2006. Int. Conf.*, vol. 3, pp. 4 pp.–1827, Feb 2006.

[2] M. Monajjemi, J. Wawerla, R. Vaughan, and G. Mori, "HRI in the sky: Creating and commanding teams of UAVs with a vision-mediated gestural interface," in *Intelligent Robots and Systems, 2013 Int. Conf. on*, pp. 617–623, Nov 2013.

[3] D. Kim, J. Lee, H.-S. Yoon, J. Kim, and J. Sohn, "Vision-based arm gesture recognition for a long-range human-robot interaction," *The Journal of Supercomputing*, vol. 65, no. 1, pp. 336–352, 2013.

[4] S. Waldherr, R. Romero, and S. Thrun, "A gesture based interface for human-robot interaction," *Autonomous Robots*, vol. 9, no. 2, pp. 151–173, 2000.

[5] C.-C. Lien and C.-L. Huang, "Model-based articulated hand motion tracking for gesture recognition," *Image and Vision Computing*, vol. 16, no. 2, pp. 121 – 134, 1998.

[6] A. Efros, A. Berg, G. Mori, and J. Malik, "Recognizing action at a distance," in *Computer Vision, 2003. Int. Conf. on*, pp. 726–733 vol.2.

[7] G. Rigoll, A. Kosmala, and S. Eickeler, "High performance real-time gesture recognition using hidden markov models," in *Gesture and Sign Language in Human-Computer Interaction* (I. Wachsmuth and M. Frhlich, eds.), vol. 1371 of *Lecture Notes in Computer Science*, pp. 69–80, Springer Berlin Heidelberg, 1998.

[8] R. Cutler and M. Turk, "View-based interpretation of real-time optical flow for gesture recognition," in *2013 Int. Conf. and Workshops on Automatic Face and Gesture Recognition (FG)*, pp. 416–416, IEEE Computer Society, 1998.

[9] X. T. et al., "Periodicity detection of local motion," in *Multimedia and Expo, 2005. Int. Conf. on*, pp. 650–653, July 2005.

[10] P.-S. Tsai, M. Shah, K. Keiter, and T. Kasparis, "Cyclic motion detection for motion based recognition," *Pattern Recognition*, vol. 27, no. 12, pp. 1591–1603, 1994.

[11] M. Allmen and C. Dyer, "Cyclic motion detection using spatiotemporal surfaces and curves," in *Pattern Recognition, 1990. Int. Conf. on*, vol. i, pp. 365–370 vol.1, Jun 1990.

[12] R. Polana and R. Nelson, "Detection and recognition of periodic, nonrigid motion," *Int. Journal of Computer Vision*, vol. 23, no. 3, pp. 261–282, 1997.

[13] R. Cutler and L. Davis, "Robust real-time periodic motion detection, analysis, and applications," *Pattern Analysis and Machine Intelligence, IEEE Trans. on*, vol. 22, pp. 781–796, Aug 2000.

[14] I. Laptev, S. Belongie, P. Perez, and J. Wills, "Periodic motion detection and segmentation via approximate sequence alignment," in *Computer Vision, 2005. Int. Conf. on*, vol. 1, pp. 816–823, Oct 2005.

[15] F. Liu and R. Picard, "Finding periodicity in space and time," in *Computer Vision, 1998. Sixth Int. Conf. on*, pp. 376–383, Jan 1998.

[16] Y. Ran, I. Weiss, Q. Zheng, and L. Davis, "Pedestrian detection via periodic motion analysis," *Int. Journal of Computer Vision*, vol. 71, no. 2, pp. 143–160, 2007.

[17] B. G. Quinn and E. J. Hannan, *The estimation and tracking of frequency*, vol. 9. Cambridge University Press, 2001.

[18] P. Borges, "Pedestrian detection based on blob motion statistics," *Circuits and Systems for Video Technology, IEEE Trans. on*, vol. 23, pp. 224–235, Feb 2013.

[19] A. Briassouli and N. Ahuja, "Extraction and analysis of multiple periodic motions in video sequences," *Pattern Analysis and Machine Intelligence, IEEE Trans. on*, vol. 29, pp. 1244–1261, July 2007.

[20] S. Seitz and C. Dyer, "View-invariant analysis of cyclic motion," *Int. Journal of Computer Vision*, vol. 25, no. 3, pp. 231–251, 1997.

[21] J. Sattar and G. Dudek, "Where is your dive buddy: tracking humans underwater using spatio-temporal features," in *Intelligent Robots and Systems, 2007. Int. Conf. on*, pp. 3654–3659, Oct 2007.

[22] M. Takahashi, K. Irie, K. Terabayashi, and K. Umeda, "Gesture recognition based on the detection of periodic motion," in *Optomechatronic Technologies, 2010 Int. Symp. on*, pp. 1–6, Oct 2010.

[23] M. Monajjemi, J. Bruce, S. A. Sadat, J. Wawerla, and R. Vaughan, "UAV, Do You See Me? establishing mutual attention between an uninstrumented human and an outdoor UAV in flight," in *Intelligent Robots and Systems, 2015 Int. Conf. on*, IEEE, 2015.

[24] F. e. a. Pedregosa, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[25] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise.," in *KDD*, vol. 96, pp. 226–231, 1996.