In *Proc. IEEE Workshop on Detection and Recognition of Events in Video* (in association with ICCV 2001)

Detecting Independently Moving Objects and Their Interactions in Georeferenced Airborne Video

J. Brian Burns
A. I. Center, SRI International
Menlo Park, CA, 94025
burns@ai.sri.com

Abstract

In airborne video, objects are tracked from a moving camera and often imaged at very low resolution. The camera movement makes it difficult to determine whether or not an object is in motion; the low-resolution imagery makes it difficult to classify the objects and their activities. When computable, the object's georeferenced trajectory contains useful information for the solution of both of these problems. In this paper, we describe a novel technique for detecting independent movement by analyzing georeferenced object motion relative to the trajectory of the camera. The method is demonstrated on over a hundred objects and parallax artifacts, and its performance is analyzed relative to difficult object behaviors and camera model errors. We also describe a new method for classifying objects and events using features of georeferenced trajectories, such as duration of acceleration, measured at key phases of the events. These features, combined with the periodicity of the image motion, are successfully used classify events in the domain of person-vehicle interactions.

1. Introduction

Key to the task of video surveillance is the detection of objects moving independently of their environment and the characterization of their type and activity. In airborne video, surveillance is in part characterized by two challenges: the constant motion of the camera, which produces spurious detections of movement on the ground, and the small image size of important objects.

Camera motion produces parallax artifacts, the appearance of independent motion in objects that are actually fixed but separated from their surroundings by some distance, such as a tree rising above the ground. This apparent movement can persist as a smooth, continuous motion for up to minutes and can lead to erroneous detection of ground activities. Figure 1 shows a frame from a video of a road through a forest, with the detected tracks of two people running across the road and four artifacts.

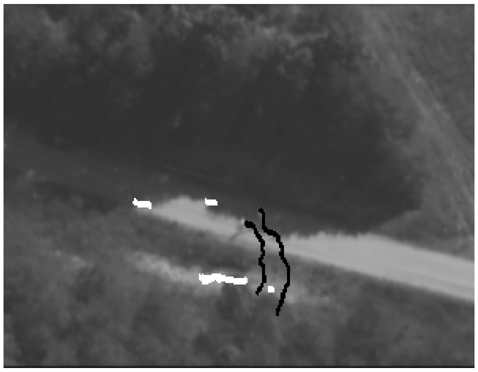


Figure 1: An example of parallax artifacts in airborne video. This is a frame showing a road cutting through woods with tracks of detected objects overlaid. People crossing are in black and artifacts generated by trees are in white.

In airborne video, moving objects of interest are often very small relative to their distance from the camera, such as a person being viewed from 2,000 feet away. Since activities can involve many objects operating over a large area, a wide field of view relative to object angular size is commonplace, often making objects only 10-15 pixels high. For example, Figure 2 shows a typical frame in the video collection studied in this paper. Because of this and the difficulty of achieving consistent motion-based image segmentation, static characteristics of the object, such as intensity pattern, shape and size [5], [7], [4], can be difficult to use for classification of object and activity. Coarse characteristics of the change in the object's image over time, such as periodicity, have also been used to understand objects [3]. However, in our domain, where the object is often traveling over highly textured ground, features of this sort, although useful, provide only a partial solution.

Reliable georeferenced video is becoming increasingly practical via calibrated camera kinematics and, even more promisingly, via precision registration of the video to reference images [14]. For the challenge of both parallax detec-

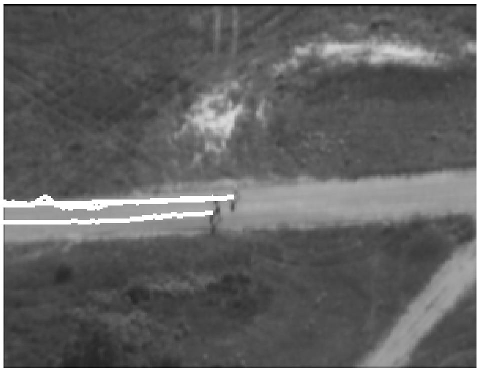


Figure 2: Synopsis of a video of people running, with detected tracks in white. The ground is basically flat, and the images of the people are about 15 pixels high.

tion and object understanding, we show that georeferenced trajectories contain very useful information that can contribute to solving these problems. For the analysis of parallax, we describe a novel method that hypothesizes a fixed object for each detected track, solves for the optimal object position given the trajectory, and then examines two features: the fit error and object height. This method was used to classify over 100 automatically detected objects and parallax artifacts with an error rate of below one percent. We also analyze the performance relative to difficult object motions and orientation errors in the camera model.

In the second part of the paper, we apply georeferenced trajectories to the classification of objects and events in the context of people interacting with vehicles. In this domain, the classification of an event is tied to the nature of the tracked objects. For example, both vehicle "parking" and "pullout" events can be observed as a tracked object stopping, followed by another object starting up nearby and a little later. The interpretation of "parking" versus "pullout" depends on which object is seen moving first: a person or a vehicle. Also, correctly classifying the objects involved depends in part on the analysis of how they are performing their actions during the event: are they stopping the way a vehicle does or more like a person? We show how characteristics of the georeferenced trajectory, such as the duration of acceleration and the achieved velocity, combined with the periodicity feature developed in [3], can be useful for simultaneously classifying objects and their interactions. In [8], object tracks are interpreted as events, but object recognition and metrical properties of the motion are not employed. In [12], metrical properties are used to cluster georeferenced trajectories and detect unusual behavior. However, classification of the objects and analysis of their interactions are not considered. Also, the measurements used are not intrinsic to the objects themselves, but are properties such as position and direction in a fixed frame. The trajectory features studied in this paper are invariant to object location and direction.

2. Detection of independently moving objects

We are interested in the detection of objects moving relative to a fixed scene in situations where the camera itself is in motion. A common method of detecting moving objects in this context is to model the image motion induced by the camera, remove this motion by warping the image with the inverse transformation (image stabilization), and classify any significant and persistent residual motion as an independently moving object [6], [1], [3], [9].

A standard practice is to assume that the scene is roughly planar (i.e., a flat ground) and model the induced image motion as a 2D affine or projective transformation. This tends to work well, even when the ground is not very flat. However, when something abruptly rises above the ground, such as a tree, telephone pole, or building, its motion is not well predicted by the ground model and it can often be detected as independently moving. Since this effect is related to motion parallax, objects of different depths having different observed motions, we will refer to the resulting false detection as a *parallax artifact*. Figure 1 shows examples of trees generating parallax artifacts and, among them, people in motion on the ground. Since we are interested in detecting and tracking people with small image size and potentially slow motion (e.g., walking), simple thresholds on image size and speed are not always suitable to properly filter out these artifacts.

The planar-plus-parallax methods [6], [10] use a more sophisticated model of image motion capable of stabilizing images of complex scenes. In these methods, a dominant planar motion is estimated as well as the lines along which the residual parallax motion is expected. (The lines intersecting the epipoles associated with the camera translation.) Given these lines, it is possible to determine which image motions are consistent with parallax and ignore them. This approach provides a self-contained and elegant solution to the problem of detecting independent motion; it is useful when camera information is otherwise unavailable and there is a sufficient amount of raised 3D structure to determine the parallax geometry. However, for uncalibrated methods such as those in [6], [10], the presence of 3D structures not in the dominant plane are required to solve for the epipolar lines [6]. When the ground is relatively flat and few fixed objects are raised above the ground, a commonly occurring situation on open roads, parking lots and fields (For example, see Figure 2.), it is not always possible to solve for the lines and the method discussed in this paper may be more useful.

We assume that the video image is registered to a ground reference frame (georeferenced) and that the altitude for every point is known to some approximation, as well as the

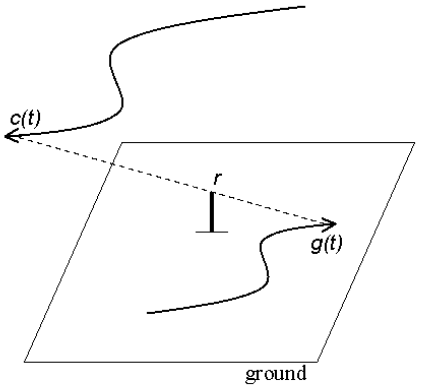


Figure 3: The geometry of parallax artifacts. The trajectory of the camera $c(t)$ sweeps out a geolocated artifact trajectory $g(t)$ at points where the line between an elevated object point r and $c(t)$ intersects the ground.

position of the camera in space. Thus, the 3D positions of a tracked object and the camera are known for each frame. For airborne video this framework is becoming increasingly practical, given progress in precision registration of video to references images [14] and the ease of localizing the camera through use of global positioning systems. Calibrated camera kinematics can also provide enough precision in georeferencing for our method and has been successfully used in the experiments presented here.

Given this framework, a relatively simple and robust method can be used to help classify a detection as an independently moving object or a parallax artifact. At any given time t, we know the ground position $g(t)$ of the detection and the position $c(t)$ of the camera (see Figure 3). If the detection is truly an artifact of the camera motion, there must be a fixed raised object r such that $g(t)$ is always in the projection ray from $c(t)$ through r, for all times t. In other words, $c(t)$, r and $g(t)$ must always be in the same line. (Note that this is mathematically analogous to a pinhole camera, where r is the "focal point" and $g(t)$ is the "image" of $c(t)$ on the ground.)

We can solve for a hypothetical r that minimizes the squared error between $g(t)$ and its estimate given r and $c(t)$, for all t, using

$$e^2 = 1/n \sum_{t=t_1}^{t_n} \left(\frac{(x_{c,t} - x_r)(z_{g,t} - z_r)}{z_{c,t} - z_r} + x_r - x_{g,t}\right)^2$$

$$+ \left(\frac{(y_{c,t} - y_r)(z_{g,t} - z_r)}{z_{c,t} - z_r} + y_r - y_{g,t}\right)^2$$

for n time samples. By eliminating the denominator, we can solve for a related least squares with a linear error:

$$e^2 = 1/n \sum_{t=t_1}^{t_n} (x_{c,t} z_{g,t} - x_{c,t} z_r - x_r z_{g,t} - x_{g,t} z_{c,t} + x_r z_{c,t} + x_{g,t} z_r)^2$$

$$+ (y_{c,t} z_{g,t} - y_{c,t} z_r - y_r z_{g,t} - y_{g,t} z_{c,t} + y_r z_{c,t} + y_{g,t} z_r)^2$$

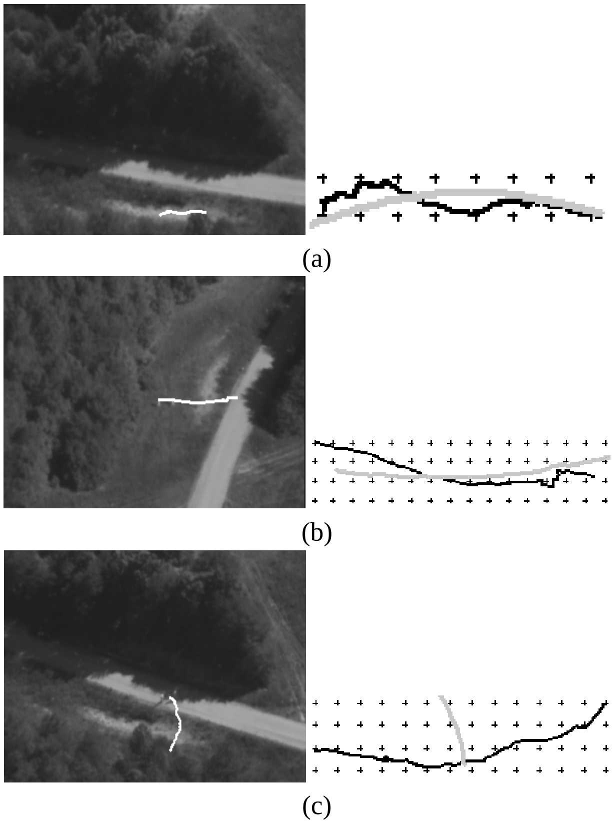

(a)
(b)
(c)

Figure 4: Examples of parallax analysis. For each, the left side shows a frame with a detected track (white) and the right side shows the geolocated track (black) and fit parallax trajectory (grey). Crosses mark 1m spacing. (a) Parallax artifact, $(e, z_r) = (0.4m, 12.1m)$. (b) Person moving in direction opposite expected for parallax, $(e, z_r) = (1.1m, -50.1m)$. (c) Person moving in roughly an orthogonal direction to parallax, $(e, z_r) = (6.2m, -2.2m)$.

In the results presented here, we solved the second equation and then applied the solution r to the first equation to compute the error e in ground units. It is important to note that we place no constraints on the 3D motion of the camera or the object when solving for r and e.

The error e and computed height z_r both tell us something about the likelihood that the detected trajectory really is a parallax artifact. Clearly if e is quite large, then the fit is not good and the trajectory is likely to be generated by something moving. If z_r is negative or very large, this also indicates independent motion. A z_r below zero (a hypothetical object below the ground) is caused by fitting to a detected trajectory moving in a direction opposite to the one predicted given parallax; a z_r that is unusually high is caused by something moving fast. Figure 4 shows examples of trajectories and their interpretation. Figure 4(a) is an ar-

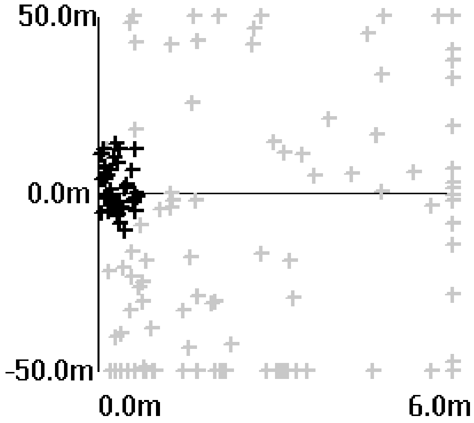


Figure 5: Distribution of trajectories in (e, z_r) space. Horizontal axis is e, and vertical is z_r. True parallax artifacts (black) are clustered near (0,0), while moving objects (grey) are more widely distributed.

tifact caused by a tree. The fit error e is $0.4m$, which is low given the observed noise in the tracking, and the hypothetical height z_r is $12.1m$, which is reasonable for a tree. In 4(b), the trajectory is from a person moving in a direction opposite to predicted parallax. It does not have a high error $(1.1m)$, but does have a height of $-50.1m$. Finally, in 4(c), the trajectory is from a person moving roughly orthogonal to predicted parallax, generating a large error of $6.2m$ relative to the predicted trajectory, which is less than $4m$ in length.

2.1. Experimental results

Figure 5 shows the distributions for parallax artifacts and independently moving objects in (e, z_r) space given a total of 142 samples (40 artifacts and 102 moving objects). This collection is composed of all tracks that were detected in a set of five video sequences and have a duration longer than 1.5s. The videos covered areas of forest, open fields and small hills, and the moving objects included people and vehicles. The movements were detected, tracked and georeferenced from airborne video using methods discussed in [1], [3], [11]. The ground motion was stabilized using an affine model, and georeferencing was principally done using calibrated camera kinematics and a single ground altitude measurement for the area.

Notice in Figure 5 that the parallax artifacts sometimes generate negative z_r values. This is probably occurring when the tree canopy, not the ground, is being stabilized. This situation can produce spurious detections down on the ground at points of high contrast such as at shadow edges. For example, observe the artifacts in the upper part of Figure 1. In spite of the fact that the ground is not always stabilized, the artifacts are clearly clustered near $(0, 0)$ in Figure 5 and the independently moving objects are much more widely distributed in both dimensions. Using thresholds that bound all the parallax samples ($-10.75 < z_r < 12.39$, $e < 0.67$), we found classification error rates of zero percent for the parallax samples and less than one percent for the moving objects. Our method and these thresholds were also used to successfully filter out parallax artifacts in the 10 videos used for the event detection experiments discussed in the final section of this paper.

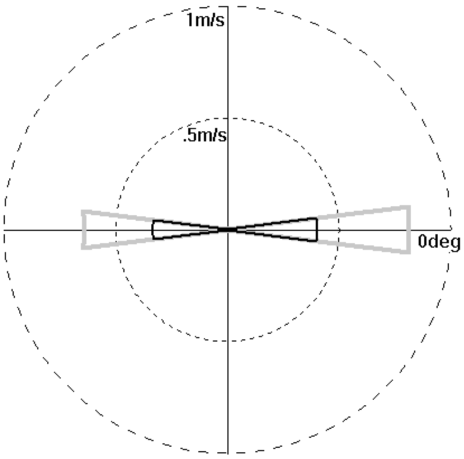


Figure 6: Polar plot of the minimum speed required for an object to be detected as independently moving, as a function of motion direction. Zero degrees is in the predicted direction of parallax. The black curve shows the minimum speeds given the (e, z_r) thresholds used in the experiments; the grey curve doubles the z_r thresholds. See text for other parameters of the simulation.

2.2. Analysis of parallax misclassification

Consider the two examples of people crossing a road shown in Figure 4(b) and (c). At what speeds and directions relative to the camera motion would their crossing of this area be mistaken for parallax artifacts and hence go undetected? The length of their traversal was approximately 15m. To simulate the effects of changing the speed and direction, we assume a straight 15m object track and a straight camera trajectory. (In situations where the object and camera trajectories have different shapes, it is harder for our method to confuse moving objects with parallax. Thus, we are simulating the worst case.) We also assume the camera is moving at $15.5m/s$ (approximately 90 miles per hour) and has a height and ground distance from the object of 500m, approximately the experimental situation above. Given these parameters, Figure 6 plots the minimum speed an object must have to avoid misclassification for every given direction of travel relative to the camera's direction of motion. In the plot, a direction of zero degrees means that the object is moving in exactly the direction expected for parallax artifacts, which is in the opposite direction of the camera's motion. The black curve plots the minimum speeds given the classification thresholds for (e, z_r) that were determined experimentally and used in the last subsection. There are two lobes, at 0 and 180 degrees, because the lower limit for parallax height z_r is a negative value of non-trivial magnitude to compensate for stabilization behavior.

The lobe widths are both 16 degrees; therefore, object directions more than 8 degrees from both 0 and 180 produce no misclassifications at any speed. Within 8 degrees of these directions, the object's trajectory begins to line up well with motion parallax and the object speed must be greater than 0.41m/s to be detected. This is a fairly low speed: moving at this pace, an object would take 37.5 seconds (over a half a minute) to cross the road cut. Thus, for this road crossing case and the classification thresholds used in our experiments, object motion has to be quite slow and closely lined up with the expected parallax direction to be confused with parallax.

The grey curve in Figure 6 plots the minimum speeds required when the parallax height (z_r) thresholds are doubled in magnitude to [-21.5m, 24.78m] ([-70.5ft, 81.3ft]). This is considerably higher than what is required for the environment studied in this paper, but for other situations, such as higher trees, this may be necessary. For misclassification to occur in these cases, the direction must still be within 8 degrees of parallax and the speed must now be less than .82m/s (double the speed above). This speed is possible for a person, but it is fairly slow and implies a road crossing of 18.3 seconds. In conclusion, there will be objects with just the right speed and direction to appear as artifacts of the camera's motion; however, given our method, the misclassification rate approaches a minimum in many practical situations. In addition, it is often possible to track objects for extended periods. In such cases, the object will often change its motion and become differentiable from parallax.

2.3. Effect of camera pointing errors

Our method depends on a model of the camera. An important source of errors in georeferenced data is the estimate of the camera pointing angle relative to the ground. In this section, we show that our method is relatively insensitive to errors in this estimate. In our experiments, errors in camera azimuth and elevation were typically less than a degree. However, in the system reported in [14], the mechanical calibration introduced higher errors that ranged from 2 to 3 degrees. The effects that errors in this range have on the parallax features (e, z_r), given typical viewing situations and a range of elevation angles, can be estimated by simulation. In this simulation, we used a fixed object of height $z_r = 10m$ above a flat ground and a camera passing by the object on a straight path at a constant height of 500m. To best compare results across different elevation angles, for each elevation angle, the ground distance between camera and object, and the duration of the tracking, were set so that the georeferenced trajectory of the resulting parallax was 10m (when no camera errors have been introduced). Then, for each elevation angle, the parallax was synthesized and georeferenced 4 times, each time with a different type of error in the camera pointing angle estimate: elevation off by 3 or -3 degrees, and azimuth off by 3 or -3 degrees. For each of these 4 trials, e and z_r were measured and the values with the largest deviation were recorded.

The elevation angles tested ranged from 25 to 65 degrees (measured down from the horizon). The largest errors for both e and z_r occurred at 25 degrees, the pointing angle closest to the horizon. Even at this angle, errors of 3 degrees in the camera estimate produced relatively small deviations in the features. The largest parallax fit error e was 0.2m, which is well within the experimentally derived threshold of 0.67m and less than errors due to tracking. The largest deviation in the height estimate z_r was also relatively small: the object height (10m) was estimated to be 11.3m. This is an error of 13%, well within the error range that can be expected from tracking. In summary, the parallax detection method presented here is relatively insensitive to pointing angle error, which is a common source of errors in georeferenced data.

3. Detection and classification of interactions

Given the observed motions of multiple objects on the ground plane, a richer understanding of what is going on can be achieved by detecting distinct types of interactions between the objects. For example, consider the case where an object is seen moving for a while, it stops, and then soon afterward and near to where the object stopped, another object starts moving and continues for some time. There is a good chance that the two objects are interacting, but what more can we deduce given only low-resolution information about the object motions? If we knew that the first object is a person and the second is a vehicle, then it would be much easier to conclude that we were observing a vehicle "pullout" event, rather than a "parking" event. In this section, we study, for the domain of person-vehicle interactions, the classification of both the interaction and the objects involved by analyzing how the interaction is executed (i.e., the motions of the objects).

We consider interactions involving a pair of objects, and use the relative position and timing of their detection, as well as the low-resolution motion characteristics of the two objects, to detect and classify the event. Two basic types of low-resolution motion information are available: the trajectory of each object relative to the ground and the coarse

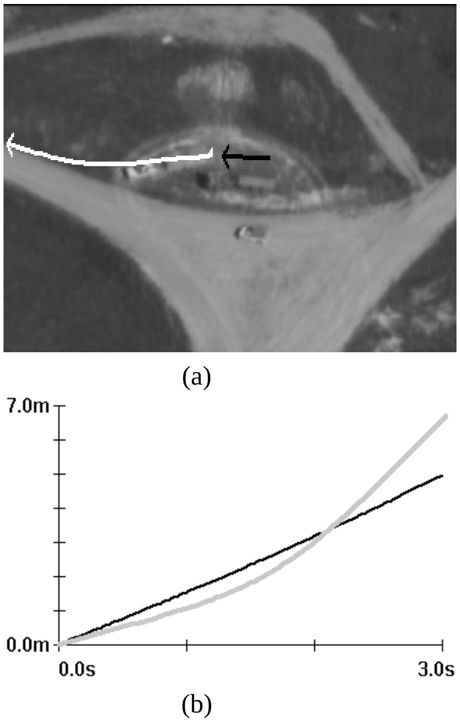


Figure 7: Example of a “pullout” event. (a) A video frame with tracks of a person in black, and the vehicle he gets into and drives away in white. The tracks have been georeferenced and then projected onto a frame taken when the person had already stopped and the vehicle had started moving (barely visible under its white track). (b) The black curve is the distance the person accumulated over time for the *last* 3-second interval of his movement, and the grey curve is distance accumulated for the vehicle (after the person stopped) for the *first* 3 seconds of its movement.

changes in each objects image as it moves over time. Both say something about how the object is performing its actions.

Picture a vehicle coming into view and stopping; now picture a person doing the same thing. Because of its mass and how it is driven, a vehicle will typically decelerate over a longer period of time than a person. In fact, since we need a certain minimum speed to detect the object in the first place and a minimum period of time to measure the accelerations, the deceleration of a person stopping in airborne video is often not observable. The situation is also the same for accelerations of vehicles and people starting their movements from a full stop. Figure 7(a) shows a frame from a vehicle pullout event. The black track represents a person approaching and then stopping to get into a vehicle, and the white track represents the movement, sometime later, of the vehicle driving away. The motion characteristics of the two objects can be appreciated by plotting the georeferenced distance accumulated over time for the last 3 seconds of the person (See Figure 7(b), black curve) and the first 3 seconds of the vehicle (grey curve). Observe the relatively straight curve of the person walking and the more apparent acceleration in the vehicle starting up.

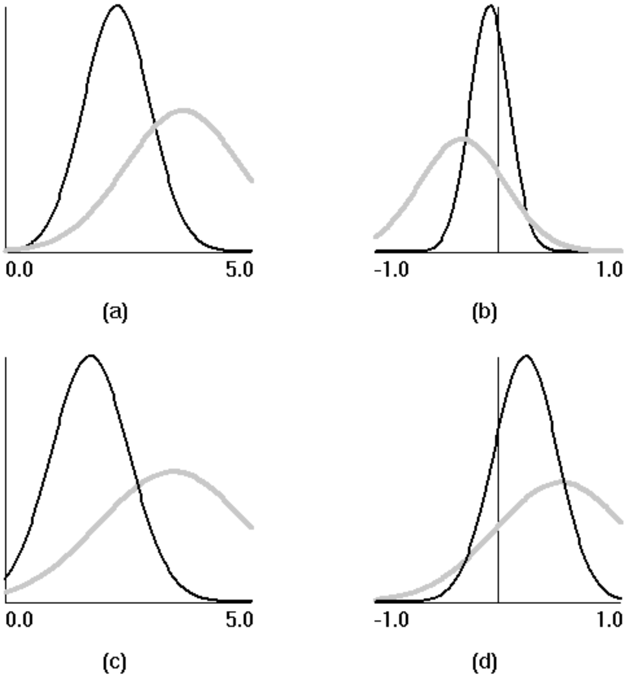
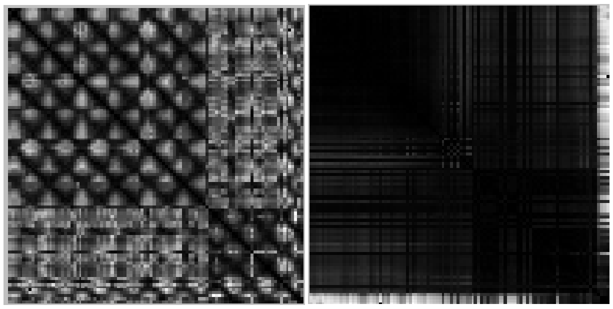


Figure 8: Distributions of velocity and acceleration features for people (black) and vehicles (grey) modeled as Gaussians. (a) Peak velocity for the first 5 seconds. (b) Acceleration for the same period. (c) Peak velocity for the last 5 seconds. (d) Acceleration for the same period.

If we approximate the acceleration of the vehicle as being relatively constant over an interval of size T, and the person’s acceleration as being more abrupt and then flat thereafter in the same interval, then we can differentiate the two objects by fitting the distance over time in the interval with a quadratic and using two times the second-order term for a measure of acceleration a. For the case of the vehicle, a should be large if the average acceleration in the interval is large (which is the difference between beginning and ending speeds). But for the person’s case, the model should not fit as well, and the acceleration feature a should be much lower than expected given the beginning and ending speeds. Basically, the quadratic will fit the roughly straight curve of the person’s accumulated distance with an almost straight line. In the example in Figure 7, T is $3s$, a for the vehicle is $1.38m/s^2$, and a for the person is $0.14m/s^2$, which actually has the wrong sign and is probably a product of detection noise.

Vehicles and people also tend to attain different speeds, which can be estimated by evaluating the above quadratic at time T before (after) the object stops (starts). For this ex-



(a) (b)

Figure 9: Correlation between every pair of images of a tracked object within a 100 frame interval . At every (row, column) position there is a correlation value for a different pair of images; dark represents high correlation. (a) Correlation matrix for a person. (b) Matrix for a vehicle.

ample, the velocity v is $4.1m/s$ for the vehicle and $1.4m/s$ for the person. Figure 8 shows the distributions of v and a for 41 event samples and an interval T of $5s$, where the distributions are modeled as Gaussians. In each case, there is a significant spread in the distribution, but enough distance between the vehicle and people means for the four features to be useful in classifying the events. Some of the spread is due to incomplete detection and tracking of the objects: the period of acceleration or peak velocity is sometimes missed. Observe that the acceleration spread is greater for vehicles than people; this is probably because the vehicle's large size and relatively low texture in the top surface creates multiple detections whose localization about the vehicle image is unstable. Improved vehicle tracking, such is in the work of [13], should help the acceleration estimates significantly. It is important to note also that georeferenced trajectories have the advantage over ungeoreferenced image tracks of being invariant to viewing conditions.

In [3], an aspect of the coarse image variation, periodicity, is used to classify walking people versus vehicles. We exploit this feature, as well as the aspects of the georeferenced trajectory discussed above. To compute periodicity, the object's image in each frame is correlated with its image in all other frames within some time interval of size N, typically around 100 frames. The NxN correlations are placed in a matrix, where each (row, column) location represents a different image pair. Figure 9(a) shows the regular pattern of correlation variation as a person moves through the different phases of walking, while 9(b) shows the lack of variation typical for vehicles. The periodicity feature is a function of the clarity and regularity of this pattern, measured using various signal processing steps discussed in [3]. In our system, it is scaled so that 0.0 is no evidence of the walking pattern, and 1.0 is the strongest evidence of the pattern.

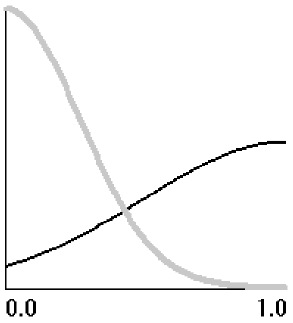


Figure 10: Distribution of periodicity values for people (black) and vehicles (grey) modeled as half-Gaussians.

Periodicity is a useful feature to exploit for our discrimination tasks. However, it is often inconsistent in situations where the background is highly textured, the object size is very small or the object-background contrast is poor. In our domain, all three of these can happen. In [2], Cutler reports the results of applying his measure to objects from the video sequences studied in this paper. He found a false negative rate of 73% for detecting people when setting the periodicity threshold to reject all vehicles, and concludes that the background intensity variation has a strong impact on the results. In this study, we enhanced his feature in some ways and found improved performance. Since the most common failure of the approach is not observing the periodicity clearly when a person is walking, we compute periodicity every 50 frames and use the maximum score across the whole trajectory as the periodicity for the object. We also use filtered (smoothed) object tracks to better localize the image for correlation. With these enhancements, a person is misclassified about 50% of the time when the periodicity threshold is set as above. Figure 10 shows the periodicity distribution for people and vehicles modeled as half- Gaussians.

Given the features of the trajectory (velocity and acceleration) and the image variation (periodicity) for the two objects involved in an interaction, we have six features that can be used together to discriminate “pullout” events from “parking” events.

3.1. Object interaction experiments

In the experiments presented here, object interactions were detected and classified based on space and time relationships between the detected objects, and the six motion features.

For each video sequence tested, we first removed the detected parallax artifacts in the manner discussed. Since moving vehicles were often fragmented into as many as 10 separate and simultaneous regions of motion, regions moving in a close cluster were automaticaly grouped and treated as a unit. The motion features of all the trajectories in a group were averaged together to further tighten the distribution of their values.

Interactions were then detected between trajectory groups. An interaction event was detected for every pair of groups that has (1) the correct time relationship, (2) spatial proximity and (3) high enough probability of being of type "pullout" or "parking". To have the correct time, all the detections of the second group must happen strictly after all the detections of the first. To be close enough, the minimum distance between ending points of the first group and beginning points of the second group must be within 8m. (This tolerance largely reflects incomplete tracking.) Finally, the pair must be classified as either "pullout" or "parking" with a probability of at least 0.8. The probabilities were estimated by applying Bayes' formula to the density functions of the six features shown above, assuming independence and that the two events are equally likely.

In a collection of 10 videos containing a total of 30 events (15 of each type), all but one of the actual events were correctly detected and classified, with no false detections. The single missed pullout event was due to a failure to detect the person approaching the vehicle. The example in Figure 7 was correctly classified as "pullout" with a probability of 0.95.

4. Conclusions

In this paper, we demonstrate that georeferenced trajectories contain useful information that can help discriminate independent motion from parallax artifacts and classify objects and their interactions. Our method of detecting parallax artifacts was experimentally demonstrated on over one hundred trajectories. It was shown by simulation that the method is able to detect independent motion even when the motion is relatively slow and in a direction very close to that expected for parallax. It was also shown that the discrimination rate is insensitive to camera pointing angle errors that are within practical ranges. Our method was developed and tested using airborne video; however, we also expect it to be appropriate whenever the camera motion is known relative to an observed surface.

In the second part of the paper, we show how objects and events can be classified using features of the georeferenced trajectories, such as the duration of acceleration measured during critical parts of the events. In our experiments, we exploited these trajectory features, as well as the periodicity of image motion, to classify events in the domain of person-vehicle interactions. Our approach should also be useful in other contexts, such as cameras mounted on buildings, as in [12]. In future work, we plan to combine the features with additional ones, including object shape and size, and perhaps the shape of the trajectories, to enhance discrimination of objects and events.

Acknowledgments

Bob Bolles, Marty Fischler, Tuan Luong and Ross Cutler were very helpful in many parts of this work. The work was supported by DARPA under contract DAAB07-98-C-J019.

References

[1] R. Cutler, C. Shekhar, J. Burns, R. Chellappa, R. Bolles, and L. Davis, "Monitoring human and vehicle activities using airborne video," in *Proc. 28th Applied Imagery Pattern Recognition Workshop (AIPR)*, 1999.

[2] R. Cutler, personal communication, July 7th 1999.

[3] R. Cutler and L. Davis, "Robust Real-Time Periodic Motion Detection, Analysis and Applications," in *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(8): 781–796, August 2000.

[4] H. Fujiyoshi and A. Lipton, "Real-Time Moving Object Classification by Image Skeletonization," in *Proc. DARPA Image Understanding Workshop*, pp. 137-144, November 1998.

[5] I. Haritaoglu, D. Harwood, and L. Davis, "W^4: Real-Time Surveillance of People and Their Activities," in *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(8): 809–830, August 2000.

[6] M Irani and P. Anandan, "A Unified Approach to Moving Object Detection in 2D and 3D Scenes," in *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20: 557–589, 1998.

[7] A. Lipton, H. Fujiyoshi and R. Patil, "Moving Target Classification and Tracking from Real-Time Video," in *Proc. DARPA Image Understanding Workshop*, pp. 129-136, November 1998.

[8] T. Olson and F. Brill, "Moving Object Detection and Event Recognition Algorithms for Smart Cameras," in *Proc. DARPA Image Understanding Workshop*, pp. 159-175, 1997.

[9] R. Pless, T. Brodsky and Y. Aloimonos, "Detecting Independent Motion: The Statistics of Temporal Continuity," in *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(8): 768–773, August 2000.

[10] H. Sawhney, Y. Guo, J. Asmuth, R. Kumar, "Independent Motion Detection in 3D Scenes," in *Proc. IEEE Inter. Conf. on Computer Vision*, pp. 612-619, September 1999.

[11] C. Shekhar, "Semi-automatic Video-to-site Registration for Aerial Monitoring," in *Proc. ICPR 2000*, pp. 736-739, 2000.

[12] C. Stauffer and W. E. L. Grimson,"Learning Patterns of Activity Using Real-Time Tracking," in *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(8): 809–830, August 2000.

[13] H. Tao, H. Sawhney, R. Kumar, "Dynamic Layer Representation with Applications to Tracking, " in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 134–141, 2000.

[14] L. Wixson, J. Eledath, M. Hansen, R. Mandelbaum, and D. Mishra, "Image Alignment For Precise Camera Fixation and Aim," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 594–600, 1998.