DETECTION OF CHANGES IN SURVEILLANCE VIDEOS

Longin Jan Latecki, Xiangdong Wen, and Nilesh Ghubade

CIS Dept.
Temple University
Philadelphia, PA 19122
latecki@temple.edu

Dept. of Mathematics
Temple University
Philadelphia, PA 19122
wen@math.temple.edu

CIS Dept.
Temple University
Philadelphia, PA 19122
nileshg@temple.edu

ABSTRACT

In this paper we provide theoretical and experimental results that dimension of video trajectories is a useful tool to access mid-level content of videos, like, appearance or disappearance of an object, and changes in velocity and in direction of moving objects. Moreover, the amount of changes is proportional to the size of objects involved and their speed. All this is achieved by a robust technique of dimensionality computation of video trajectories based on eigenvalues.

1. INTRODUCTION

A first step in our video analysis is to map a video sequence to a polygonal trajectory. We obtain a polygonal video trajectory in an Euclidean space by mapping each frame to a feature vector and joining the vectors representing consecutive frames by line segments.

The second step is geometric analysis of obtained trajectories. The goal of this analysis is to link the properties of video trajectories to the content of videos.

Our approach can be classified as geometric analysis of video trajectories in the feature space. This direction was started by DeMenthon at al. [1], who used it for key frame extraction. It was further advanced for key frame extraction in [2, 4, 5]. Geometric analysis of video trajectory was first applied to detect unusual events in videos in Latecki and de Wildt [6]. All mentioned techniques are based on an interpolation of the original video trajectory with a polygonal curve with significantly fewer vertices, and differ by interpolation techniques used. In this paper we propose a new technique of geometric analysis of video trajectories that is based on dimension analysis.

In order to be able to capture object motions, we use image histograms extended by centroids of histogram bins as feature vectors that represent images. This representation was proposed in [1].

For a given digital image, we compute a histogram with k bins for each color component, e.g., in YUV or RGB color space. The feature vector f_t for a frame number t consists of the following values for each color bin i in each color component: b_x^i, b_y^i (the x and y coordinates of the centroid of the pixels belonging to bin i), and $b_\#^i$ (the pixel count in bin i):

$$f_t = (b_x^1, b_y^1, b_\#^1, ..., b_x^{3k}, b_y^{3k}, b_\#^{3k})_t.$$

The pixel count $b_\#^i$ is relative to the total number of pixels in the image, i.e., it is a value between 0 and 1. Also the x and y coordinates of the centroid are scaled to range between 0 and 1 by dividing them through the x and y image size, correspondingly. For example, in [6] we used 8 bins; this led to feature vectors with 72 coordinates plus t as 73rd coordinate.

The video trajectory can be interpreted as sequence of vectors $v_t = f_t - f_{t-1}$ that join two consecutive feature vectors. We define the dimension of video trajectory to be the dimension of the smallest linear subspace containing the trajectory. It can be computed as the rank of the video trajectory matrix, whose rows are the video trajectory vectors.

However, the computation of the rank is highly sensitive to numeric errors. Thus, a robust way of rank computation is needed. Such a method is provided in Section 2. The main idea is that the rank can be defined as the number of nonzero singular values. Because of the numerical errors, we can not get the exact zero singular values, but we can view singular values with small Euclidean norm as zeros. This idea was introduces in Seitz and Dyer [8]. In the paper by Rao et al. [9], the same method is used to compute the matching error of two action trajectories.

The question arises, however, how do we link the dimension of video trajectories to the events in videos. This is based on our theoretical results in [7], and the results stated in the appendix of this paper. We prove in Theorem 2 that the video trajectory is planar when a single object (a set of pixels) is instantly visible for any kind of motion of this object. This means that the rank of the video trajectory matrix is less than or equal to two. This is a surprising result, since

TO APPEAR IN IEEE INT. CONF. ON ADVANCED VIDEO AND SIGNAL BASED SURVEILLANCE, MIAMI, JULY 2003.

video trajectory can easily reside in a several hundred dimensional space (which is the number of used features). We also prove that the dimension of video trajectory increases at least by one and at most by three if a new object appears and is moving, Theorem 3. Consequently, if the rank of part of the video trajectory is greater than two, then we can tell that a new object appeared (disappeared) in the corresponding part of the video. A disappearance of an object is equivalent to an appearance of a new object with a background texture.

The fact that we are able to detect changes in object direction and speed is based on the following results proved in [7]. Consider a video in which an object is moving on a constant background. If this object is visible in all frames and it is moving with a constant speed on a linear trajectory, then the feature vectors define a straight line in the feature space. Consequently, a constant motion of a single object along a linear trajectory in the video results in a linear video trajectory in the feature space, i.e., the dimension of the video trajectory is one. Observe that over a short period of time, the motion trajectory in the image plane of many real objects (like humans and vehicles) can be well approximated as linear trajectory. Further we showed in [7] that if a moving object changes direction or speed, the dimension of the video trajectory will increase. It can only increase by one, since, as mentioned above, when we have one object moving the dimension is not greater than two. Consequently, if a single object is moving, its change of direction or speed should be easily detectable, since the dimension of the video trajectory doubles.

Further, we proved [7] that if n objects are moving with constant speeds on a linear trajectory, then the video trajectory is a single straight line in the feature space. Thus, even with n object moving, we are able to detect changes in direction or speed of each of them as discussed above. However, this also shows a limitation of our system, we will not be able to tell which object changed its direction or speed. We are only able to tell that one of the objects did. On the other hand, the fact that we do not try to track or distinguish the objects, may be the main reason for the robust performance of our system.

We assume for all our theorems that the video background is uniform: all background pixels belong to the same color bin for each color component (i.e., background pixels contribute to at most one color bin for each color component). This is not a restriction for practical applications, since there exist several techniques for background learning and background stubstraction (e.g., see Chapters 14.3.11 and 16.2.5 in Forsyth and Ponce [3]). Further, our systematic tests with real surveillance videos show that background subtraction is not necessary for our system to perform robustly. In section 4 we present sever experimental results with real surveillance videos that demonstrate a robust performance of our system.

2. THE ERR OF THE RANK OF A MATRIX

In the paper by Seitz and Dyer [8] the distance measure

$$dist = \sqrt{\frac{1}{mn}\sum_{i=4}^{n}\sigma_i^2}$$

is used, where m, n is the size of the matrix M_Γ, $\sigma_i, i = 1, \cdots, n$ are the sorted singular values of the matrix based on their Euclidean norm, to compute the match error of a set of images Γ given their measurement matrix Γ. The match error is defined by:

$$dist_A(\Gamma) = \min\{||E||_{rms}; rank(M_\Gamma + E) \leq 3\},$$

where $||E||_{rms}$ is the root-mean-squared norm of the matrix E defined by

$$||E||_{rms} = \sqrt{\frac{1}{mn}\sum_{i,j}E_{ij}^2}.$$

Seitz and Dyer prove that

$$dist_A(\Gamma) = \sqrt{\frac{1}{mn}\sum_{i=4}^{n}\sigma_i^2}.$$

In the paper by Rao et al. [9], the same method is used to compute the matching error of two action trajectories. Because their match matrix M is 4 by n, and it has only 4 singular values, they use the distance measure as:

$$dist = |\sigma_4|$$

Based on these ideas, we define the error of rank k of a matrix M as:

$$err_k(M) = \min\{||E||_{rms}, rank(M + E) \leq k\}.$$

The larger $err_k(M)$, the larger is the difference of matrix M from a matrix of dimension k or smaller. The following theorem generalizes the result stated in Seitz and Dyer [8] for $k = 3$ to any k.

Theorem 1

$$err_k(M) = \sqrt{\sum_{i=k+1}^{n}\sigma_i^2} \tag{1}$$

where m, n is the size of the matrix M, $\sigma_i, i = 1, \cdots, n$ are the sorted singular values of the matrix based on their Euclidean norm.

Proof: Singular values decomposition gives

$$M = U\Sigma V,$$

where U and V are orthogonal matrices and the singular values sorted in descending order according to their Euclidean norm, appear along the diagonal of Σ.

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \sigma_3 & & \\ & & & \dots & \\ & & & & \sigma_n \end{bmatrix}$$

Since $\Sigma = \Sigma_1 + \Sigma_2$, where

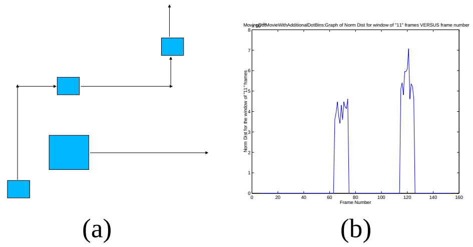
$$\Sigma_1 = \begin{bmatrix} \sigma_1 & & & & & & \\ & \sigma_2 & & & & & \\ & & \dots & & & & \\ & & & \sigma_k & & & \\ & & & & 0 & & \\ & & & & & \dots & \\ & & & & & & 0 \end{bmatrix},$$

$$\Sigma_2 = \begin{bmatrix} 0 & & & & & & \\ & \dots & & & & & \\ & & 0 & & & & \\ & & & \sigma_{k+1} & & & \\ & & & & \sigma_{k+2} & & \\ & & & & & \dots & \\ & & & & & & \sigma_n \end{bmatrix},$$

we obtain $M = U\Sigma_1 V + U\Sigma_2 V$. Since $E = -U\Sigma_2 V$ is minimal by Theorem 6.7 in [10], $U\Sigma_1 V$ is the optimal rank-k approximation of M, and $U\Sigma_2 V$ is the error matrix. Because $||U|| = ||V|| = 1$, the result follows. ■

3. DETECTION OF CHANGES

As mentioned in the introduction, our theoretical results show that the rank of a video trajectory in a feature space changes whenever a new object appears or an existing object disappears from a camera view field. In particular, if the rank is less than or equal to 2, then there are no changes in visible objects in the video, but if the rank is greater than 2, then there are changes in visible objects. This suggests to use $err_2 = \sqrt{\sum_{i=3}^{n} \sigma_i^2}$, since it tells us how much the trajectory matrix differs form the matrix of rank less than or equal to 2. In order to identify parts of videos with significant changes, we compute for the trajectory matrices in the window W_{11} of 11 consecutive frames. This way we compute a 1D function that assigns err_2 to the mid frame of each window. Significant maxima of this function identify unusual events in videos in which either a new object appears or an existing object disappears from a camera view field. In all our experiments, some of which are presented in the next section, we detected all significant changes.



(a) (b)

Fig. 1. (a) Motion trajectories in the image plane for the moving blocks movie. (b) A 1D function computed based on err_2.

4. EXPERIMENTAL RESULTS

We performed a series of systematic experiments with varying kinds of real and synthetic surveillance video clips to verify the performance of our technique.

We begin with a simple ground-truth movie. Figure 1(a) shows the path traversed by a small block and an additional larger block appearing in between of its travel. The movie consists of 160 frames. The additional block appears at frame 70 and moves in the east direction until it disappears at frame 121. In Figure 1(b), we can see two clear peaks: the first peak is at the frame number 70 when the additional block appears and the second peak is at the frame number 121 when the additional block disappears. Notice that we cannot clearly detect the turns of the small block.

Now we concentrate on several real videos. Video *security7.mpg*[1] lasts for 15 seconds and has 387 frames, ca. 25 fps. It is a low resolution video (166x112). First a hand-held camera points at a closed door in an empty room. At about frame 170, a man opens the door and enters the room. He moves around and disappears from the camera view field to the right at about frame 250. The rank err_2 computed for some time intervals is graphed in Figure 2. The first peak is at the frame number 169 when the man enters the view of camera and the second peak is at the frame number 253 when he leaves the view of camera (see Figure 3).

The results for another video clip, *Mov3.mpg,*[2] demonstrate that our approach is context-dependent. While in the previous video, we needed to detect a moving person, the goal here is to distinguish between arm movements and small movements of the upper body. Movie *Mov3.mpg* lasts for 15 seconds and has 386 320x240 frames, ca. 25 fps. A

[1] It can be viewed on www.cis.temple.edu/ ˜ latecki/Movies.

[2] It can be viewed on www.cis.temple.edu/ ˜ latecki/Movies

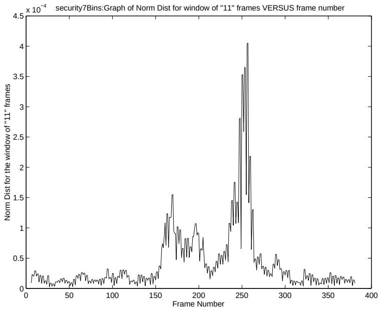


Fig. 2. A 1D function computed based on dimension err_2 of the video trajectory for *security7.avi* movie.

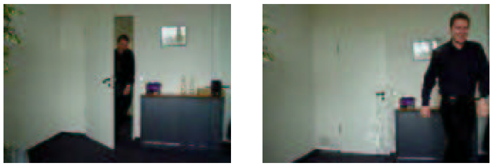


Fig. 3. The frames 169 and 253 indicated by two main local maxima of err_2 for *security7.avi*.

hand-held camera shows the upper part of a sitting man. The first peak in Figure 4 is when the man waves his right hand (frame 50), Figure 5, the second peak is when he again waves his right hand (frame 156), and the third peak is when he waves his left hand (frame 253).

The results for *hall_monitor.avi*[3] video clip demonstrate that our algorithm can deal with simultaneous motion of several objects. The goal here is to not only detect but also to distinguish between the motion of two different persons. The six highest peaks in Figure 6 indicate the six main events. The frames corresponding to the peaks are shown in Figure 7. These events are (listed by frame number):

35: A first man enters from a door and is seen in the camera view.

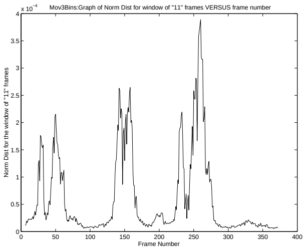


Fig. 4. (a) A 1D function computed based on dimension err_2 of the video trajectory for *Mov3.avi* movie.

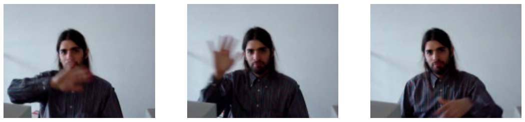


Fig. 5. The frames 50, 156, and 253 indicated by three main local maxima of err_2 for *Mov3.avi*.

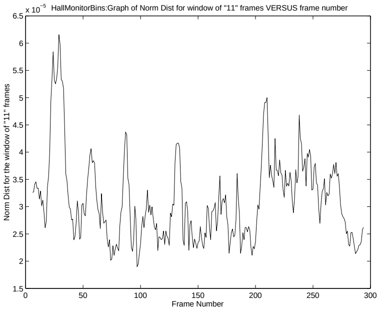


Fig. 6. (a) A 1D function err_2 computed based on dimensionality of the video trajectory for *Hall_Monitor.avi* movie.

54: He is moving away from the camera and two new objects became visible: small table and man's pants.

92: He stops and bends to keep a bag, a second man enters the view.

136: The whole body of the second man is visible.

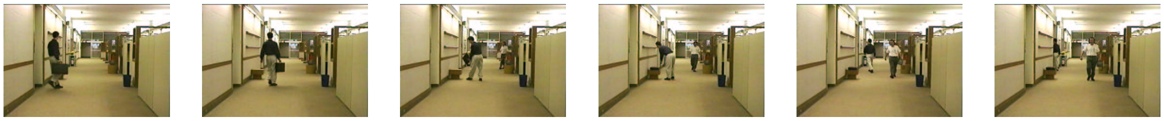
215: The first man begins to exit through the second door on the left, the second man has moved closer to the camera.

241: The first man disappears behind the door.

When computing video trajectory dimensionality to detect unusual events on movie *CameraAtLightSignal.avi* [4] we obtained graph of the dimensionality change function shown at Figure 8(a). The two main peaks of this function are at frames 23 and 85. They correspond to the two unusual events in this video–*complete* appearance in the camera view field of the first (black) and of the second (white) car. Other local maxima around the two main peaks corresponds to the variable visibility of different parts of the two cars. Observe that between the frames 30 and 60 the considered function has no significant variations since there are no unusual events on the scene (the first car moves from right to left while the second car does not appear yet).

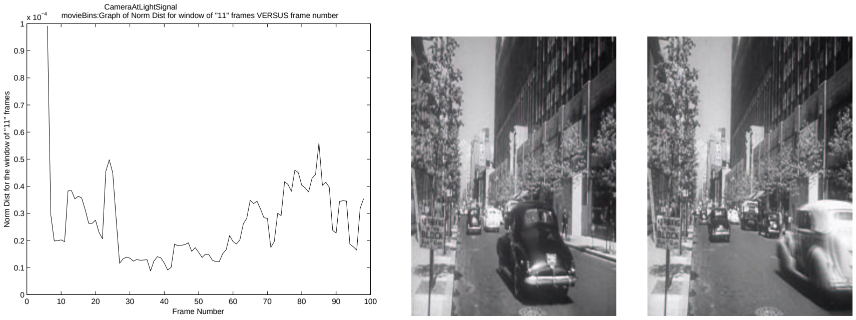
[3] It can be viewed on www.cis.temple.edu/ ˜ latecki/Movies.

[4] http://www.cis.temple.edu/ ˜ latecki/Movies



(frame 35) (frame 54) (frame 92) (frame 136) (frame 215) (frame 241)

Fig. 7. The frames indicated by six highest local maxima in Figure 6 for movie *Hall_Monitor.avi*.



(a) (b) (c)

Fig. 8. (a) A 1D function computed based on err_2 of the video trajectory for *CameraAtLightSignal.avi* movie. The frames indicated by two main local maxima of this function are shown in (b) frame 23, and (c) frame 85.

5. CONCLUSIONS

Our theoretical results lead to useful conclusions to detect changes in video by analysis of video trajectories in the feature space:

- A change in the number of visible objects implies a change in the dimension of the video trajectory by 1, 2, or 3.
- Linear motion in the image plane leads to a linear trajectory.
- A change in motion direction or speed of an object increases the dimension of video trajectory by 1.

We can detect this changes by computing rank of the trajectory vectors. Since the classical way of rank computation is very noise sensitive, we needed to employ a more stable method.

We defined the error err_k of rank k of a matrix M in formula (1). When err_k is small, it means that the rank of matrix M is less or equal to k, and if err_k is big, we can tell that the rank is bigger than k. In this paper we compute err_2 for matrices representing parts of video trajectories. The value of err_2 tells us the amount of changes of in the rank of matrices that are directly related to the changes in the video. In the experiments, we use $k = 2$, because we know that the rank is less or equal to 2 for any kind of motion of a singe object which is instantly visible. If a new object appears or a visible object disappears, the rank will be larger than two. Therefore, we can use the changes of err_2 to detect unusual events in videos.

To represent video trajectories we use a simple feature space composed of color histograms extended by centrods of the color bins.

The fact that centroids of real objects are not preserved under projection does not affect our approach at all, since we neither try to recover object trajectories in the original 3D space nor trajectories of their projections. Our video trajectories are in the feature space. By analyzing these trajectories, we can detect changes in object motion that include speed and direction changes as well as appearance and disappearance of objects. Our approach is also applicable to feature spaces other than color spaces, which include motion vectors.

Appendix

Theorem 2 *If an object (a set of pixels) is moving on a uniform background, then the trajectory vectors are contained in a plane.*

Proof: We show that each entry of the feature vector could be expressed as a linear function either of x or of y, coordinates of the center of the object (or the position of the object). Because the whole object is visible in all frames, $b^i_\#$ for each bin i is then a constant e_i.

If the background pixels do not belong to the color bin i, then b_x^i, b_y^i only depend on the position of the object. Hence $b_x^i = x + b_i$ and $b_y^i = y + d_i$, where (b_i, d_i) is the vector that translates the center of the object (x, y) to the center of all object pixels in the bin i. Hence b_x^i, b_y^i are expressed as linear functions of x and y, correspondingly.

It remains to consider the bin i that contains all background pixels. We will show that b_x^i is a linear function of x, and b_y^i is a linear function of y.

Let (a, b) be the fixed coordinates of the midpoint of each image frame. The main observation is that the coordinates of the centroid of the background can be expressed as

$$(b_x^i = \frac{(N * a - n * x)}{(N - n)}, b_y^i = \frac{(N * b - n * y)}{(N - n)}), \qquad (2)$$

which are both linear functions of x and y correspondingly.

The coordinates for the feature vector could be expressed as $a_i * x + b_i, c_i * x + d_i, e_i$, where a_i, b_i, c_i are some constants. The corresponding coordinates of the trajectory vector are equal to $a_i * \Delta x, c_i * \Delta y, 0$.

Since each column of all trajectory vectors is expressed as a scalar function of either Δx or Δy, the video trajectory has the rank two. ■

As a simple consequence of theorem 2, we obtain the following corollary:

Corollary 1 *Consider a video in which n objects are moving on a uniform background. Assume that all objects are visible in all frames. Then the dimension of the trajectory is at most $2n$.*

Proof: For each feature vector, each coordinate is a linear function of either x_i or y_i, $i = 1...n$. (the position of the i'th object in the frames). ■

Theorem 3 *If a new object suddenly appears in the movie, the dimension of the trajectory increases at least by 1. If the whole object stays in the movie, then the dimension increases at most by 3.*

Proof: First we prove that the dimension increases at least by 1. Assume the background pixels belong to bin i. Then $b_\#^i$ changes from one constant to another constant, and consequently, $\Delta b_\#^i \neq 0$ in the difference vector frame k minus frame $k - 1$, whereas all differences before frame $k - 1$ are equal to zero. Therefore, the dimension increases at least by 1.

Now we prove it increases at most by 3: Assume originally there are N objects. Then the dimension is $2N$. Assume that a new object appears at frame k. If we drop the difference vector between k and $k-1$, the rest of the trajectory vectors could be seen as the trajectory vectors for N+1 objects moving in the movie. They are then in a subspace with dimension $2(N + 1)$. Since the difference vector between k and $k - 1$ will increase the dimension by 1. We obtain that the dimension increases at most by 3. ■

6. REFERENCES

[1] D.F. DeMenthon, V.M. Kobla, and D. Doermann. Video Summarization by Curve Simplification. *Proc. ACM Multimedia*, pp. 211-218, 1998.

[2] D.F. DeMenthon, L.J. Latecki, A. Rosenfeld, and M. Vuilleumier Stückelberg. Relevance Ranking and Smart Fast-Forward of Video Data by Polygon Simplification. *Proc. Int. Conf. on Visual Information Systems*, pp. 49-61, 2000.

[3] D. Forsyth and J. Ponce. *Computer Vision. A Modern Approach*. Prentice Hall, 2003.

[4] L. J. Latecki, D. DeMenthon, and A. Rosenfeld. Automatic Extraction of Relevant Frames from Videos by Polygon Simplification. *Proc. German Conf. on Pattern Recognition (DAGM)*, pp. 412-419, 2000.

[5] L. J. Latecki, D. de Wildt, and J. Hu. Extraction of Key Frames from Videos by Optimal Color Composition Matching and Polygon Simplification. *Proc. Multimedia Signal Processing*, 2001.

[6] L. J. Latecki and D. de Wildt. Automatic Recognition of Unpredictable Events in Videos. *Proc. Int. Conf. on Pattern Recognition*, Vol. 2, 2002.

[7] L.J. Latecki, X. Wen, and N. Ghubade. What can we tell about object motion from its video trajectory. To appear.

[8] S. M. Seitz and C. R. Dyer. View-invariant analysis of cyclic motion. *Int. J. of Computer Vision* 16:147-182, 1997.

[9] C. Rao, A. Yilmaz, and M.Shah. View-Invariant Representation and Recognition of actions. *International Journal of Computer Vision* 50(2), 203-226, 2002.

[10] G. W. Stewart. *Introduction to Matrix Computations*. Academic Press,New York,NY,1973