Automatic Visual Recognition of Armed Robbery

Jaime Dever, Niels da Vitoria Lobo, and Mubarak Shah
Computer Science, University of Central Florida, Orlando, FL 32816
E-mail: {jdever,niels,shah}@cs.ucf.edu

Abstract
We propose a method by which to analyze silhouettes and recognize a classic holdup position of armed robbery. In such a situation, one actor levels his or her arm while another actor raises his or her arm(s) into the air. We attempt this by first segmenting the skeleton of the silhouette into separate pieces of the body, then identifying the positions of the arms.

1. Introduction and Related Work

Video surveillance has long been used in an attempt to prevent crimes by providing a ready means of identifying the perpetrator and ensuring that he or she is held accountable. Useful as this may be, it remains a passive method of crime control. Computer recognition of activities in such situations could thrust video surveillance into an active role by allowing police to be alerted automatically, hopefully in time to prevent loss of life or property.

Recognition of human activities is already the subject of much research. In one such project involving surveillance of parking lots[3], characterizations were based solely on the motion of the objects in the image. The movements were then compared to a set of possible events.

In another project[1], the trajectories of a set of objects were compared against models for specific activities. The actors themselves were not investigated; they were only tracked and their vectors analyzed. For example, football players were viewed from above and their motions were compared against a database of plays.

None of these holistic approaches will identify actions as specific as levelling an arm at another actor that raises its arm(s) into the air in response. In our approach, we investigate the individual motions and silhouette appearance more closely. We propose to analyze the skeleton[4], or medial lines, of the silhouette in order to identify the position of the arms and recognize the classic hold-up positions of armed robbery.

2. Overview of Algorithm

The steps in our algorithm are:
1. Silhouette Extraction: The people in the scenario must first be identified. This is accomplished through background subtraction together with connected components[4]. To achieve a smooth silhouette for skeleton calculation, this is followed by dilation and erosion[4].

2. Skeleton Segmentation: First, the skeleton is calculated. Points of interest (POIs) are then determined and used to find the individual segments of the skeleton.

3. Segment Identification: Position, length, and slope are then used to identify the torso, legs, arms, and head, if they are present.

4. Arm Analysis: Finally, the slopes of the arms are used to determine whether a possible armed robbery exists in the frame or sequence.

3. Silhouette Extraction

Simple identification in frames and tracking through a sequence is accomplished using Stauffer’s background subtraction [5] (Figure 1) and connected components. In this implementation, the initial background must be known and the camera must be fixed. The connected components phase associates all canny[2] edge pixels (Figure 1) of the foreground with other edge pixels in the foreground within a 20-pixel radius. Edges are used to simplify size thresholding of blobs and limit interference of shadows, while a large window is used to prevent severance of the head from the rest of the silhouette. The remaining foreground pixels are grouped by connected components using the same radius, within a bounding box determined by the extremes of the grouped edges. In subsequent frames, the previous location of the blob is used for the bounding box rather than the edges, after allowing for some movement. This helps ensure that previously acquired silhouettes will not be culled unnecessarily due to size restraints. Dilation and then erosion
are both completed using a window of 5 pixels, followed by a more selective erosion in which pixels with too few neighbors are erased. (Figure 1) These actions help smooth the silhouette for skeleton calculation. Only a basic silhouette is needed for the next steps, as the skeleton analysis is intended to compensate for normal errors in the extraction.

Figure 1. Preprocessing (clockwise) (a) original image. (b) background subtraction. (c) canny edges. (d) silhouette.

4. Skeleton Segmentation

The silhouette’s skeleton is calculated by eroding the edges of the region until only curves with width of one pixel remain. Then, POIs are identified. These are determined by the number of neighboring pixels in the skeleton: one neighbor indicates an endpoint, while three or more indicate an intersection. Non-POI pixels of the skeleton are then grouped by connected components to make up the individual segments of the skeleton. Their endpoints, lengths, and slopes are calculated for analysis. The actual number of segments associated with each POI is also determined and recalculated as needed when POIs and segments are added or deleted. (Figure 2)

Figure 2. Segmentation of skeleton; individual segments numbered (colored) sequentially.

5. Segment Identification

It is currently assumed that the bodies are upright. Thus, the torso is the longest segment and nearly vertical. This segment is always present; if no other segments are present, it spans the entire height of the silhouette. The nature of its POIs give clues to the presence of other segments. If a head exists, it is a small extension of the top of the torso and indicates the presence of at least one arm. The bottom point of the head corresponds to the top point of the torso, and its top point is an endpoint. Arms also have one intersection with the top of the torso and one endpoint. Legs exist when the bottom point of the torso is an intersection rather than an endpoint, and are the bottommost segments. They almost always occur in pairs, and their top point corresponds to the bottom of the torso. (Figure 3)

5.1. Error Checking

Our algorithm is designed to compensate for a number of shortcomings in the silhouette. Such faults lead to distinct errors in the skeleton.

1. Loops: Occasionally, there appear two segments that share the same endpoints. This occurs either when the legs are joined by shadow or other accident at the bottom of the silhouette, or when a hole is found in the silhouette. In the former case, the legs are separated and the missing POI is added. In the latter, the two segments are joined to be only one.
2. Spurs: Small, irrelevant segments often occur as a result of irregularities in the silhouette. In instances where one of its ends is an endpoint, the spur can simply be deleted because it is an offshoot of another segment and does not belong. The true segments wrongly separated by this error often need to be joined. In instances where the spur has another segment on both ends, the spur must be added to one of them and it must be determined whether these two segments are also one and the same. This occurs most often when the arms do not meet at the same part of the torso, so the points of these segments must be adjusted as well.

3. Links: An additional segment occasionally joins the bottoms of the legs. The cause is shadow, as with loops. In this case, the error can simply be deleted.

6. Arm Analysis

Once the arms have been identified, their positions are analyzed based on their slopes (We assume the figure to be in the first quadrant of the cartesian coordinate plane.) Using the line equation $y = mx + b$, a horizontal arm is one with $-0.4 < m < 0.4$. When an arm first becomes distinguishable from the torso, it nearly always falls within this range; this is therefore the lowest alarm level. The second level is when the victims’ arms are determined to be raised. In this case, a left arm must have a slope less than -0.8 and a right arm must have a slope greater than 0.8. The highest level of alarm occurs when there are at least two silhouettes present and the first two alarms are both recognized.

7 Results

Figure 5 represents output from different sequences in which the actor intended to trigger the level-arm alarm. In these, the 5 body segments are correctly identified and the proper alarm is successfully reached. Figure 6 results from different sequences in which the actor attempted to trigger
the raised-arm alarm. Here, the 5 body segments are correctly identified and the raised-arms alarms are successfully reached. In the second image, the legs are successfully separated and a spur successfully removed. Over 6 separate sequences, our algorithm successfully identified the different segments of the body in an average of 71 percent of the frames. The lowest rate was 53 percent and the highest rate was 86 percent. The slope analysis of the arms failed only when the arms were identified incorrectly.

8. Discussion and Future Work

Our algorithm analyzes the skeleton of a silhouette to determine the positions of the arms. It identifies segments of the body and reaches alarm states with a high rate of success. Although we compensate well for poor silhouettes, our algorithm remains sensitive to errors in them; the sequences with lower success rates all had silhouettes of lower quality than did the sequences with better success rates. Highly irregular blobs result in spurs, connections, and other extraneous information that can confuse the analysis. It is suggested that silhouettes yielding inordinate numbers of POIs and segments should undergo dilation or dilation and erosion, then be addressed again by the skeleton analysis algorithm. Additional work can be done with the compensation portion itself. Additional alarm criteria to narrow the identification of this situation as well as accept other situations are also sought.

References