The PETS04 Surveillance Ground-Truth Data Sets

Robert B. Fisher
School of Informatics, University of Edinburgh
rbf@inf.ed.ac.uk

Abstract

This paper summarizes the 28 video sequences available for result comparison in the PETS04 workshop. The sequences are from about 500 to 1400 frames in length, for a total of about 26500 frames. The sequences are annotated with both target position and activities by the CAVIAR research team members.

1. Introduction

This paper describes the video sequences used in the PETS04 workshop competition. The sequences are oriented about a public space surveillance task, and are ground truth labeled frame-by-frame with bounding boxes and also a semantic description of the activity in each frame. Altogether, there are 28 video sequences containing about 26500 labeled frames, grouped into 6 different activity scenaria.

The first group of videos was acquired at INRIA in July 2003. The sequences contained scripted activities by the research team members. The intended test scenaria are:

	Number of	Number of
Scenario	Sequences	Frames
Walking	3	3045
Browsing	6	6665
Collapse	4	4227
Leaving object	5	5848
Meeting	6	4135
Fighting	4	2499
Total	28	26419

However, almost all sequences also contained both an introductory activity by one of the researchers, as well as unscripted activity (usually walking or meetings by other employees at INRIA).

These sequences are publicly accessible at URL:
`homepages.inf.ed.ac.uk/rbf/CAVIARDATA1`

1.1 Ground Truth Labeling

Based on the CAVIAR activity representation model, each video frame has been labeled with a set of ground truth descriptions.

Each individual person was described by a bounding box (id, centre coordinates, width, height, orientation of main axis of individual), plus a description of his/her movement (inactive, active, walking, running). Individuals are only labeled once they start moving; otherwise they are effectively background. Based on the proposed semantics of the activity interpretation, each box is usually labeled with a role (fighter, browser, left victim, leaving group, walker, left object), is a participant in a situation (browsing, moving, inactive), which is a component of a scenario (Walking, Idleness, Browse, Collapse, Leaving object, Meeting, Fighting). Each box is labeled with some of the above labels in each frame.

The semantics of activity labeling were constrained by a finite-state model of the allowable behaviors. These are summarized in Section 2, which shows the allowable sequences of situations in a given scenario. In each scenario, the individual or group is observed in a sequence of situations determined by the finite state model for that scenario. When in a situation, the actor must fulfill a specific role linked to that situation. As well as the role, the ground truth labeling for the box has a qualitative assessment of the motion of the individual or group, *i.e.* whether they are running, walking, stationary but active (*e.g.* moving arms), or inactive.

Each video frame contains zero or more labeled individual or group boxes. The boxes are labeled with an identifier, which persists as long as the individual is visible. If a person disappears and then later reappears, then the individual obtains a new identity. If the person is obscured/occluded for only a few frames, then the same identity is maintained.

Similarly, groups of interacting individuals also are described by bounding boxes (id, centre coordinates, width, height, orientation of main axis of individual, list of component individual boxes), plus a description of the group's movement (inactive, active, moving). Based on the proposed semantics of the activity interpretation, each group box is usually labeled with a role (meeters, fighters, walkers), is a participant in a situation (fighting, moving, meeting, split up, inactive, leaving victim, leaving object), which is a component of a scenario.

The grammar of the ground truth file can be seen in appendix A. The web site will also provide the ground truth labels in XML shortly. An example of the current ground-truth entry for frame 517 of sequence LeftBag is:

```
frame    LeftBag    517    ibl
    ib   2
         210  247  55  39  10  wk
         wr  1.0   m  1.0   im  1.0
    eib
eibl    gbl    egbl    eframe
```

The description says: there is only individual box 2, with center at column 210 and row 247. The bounding box width is 55 pixels wide and 39 pixels tall, and the dominant orientation is 10 degrees. The target is walking (wk), fulfills the walker role (wr) with certainty 1.0, is in a moving situation (m) with certainty 1.0, which is part of the immobile scenario (im) with certainty 1.0.

1.2 Open issues

The labeling has highlighted some issues:

1. **Variability of the ground truth**

   Since the labeling was done by humans, there is a natural variation in both the parameters and occurrence of the labels, *e.g.* the positions and sizes of the bounding boxes, or when the box or activity starts. Knowing the range of human variation will help with comparison to automatic calculations of the statistics.

   To help assess this question, one of the datasets has three labelings by different individuals. As the statistics package is still being developed, we do not yet have data on the variation.

2. **Nature of the behaviour labeling**

   We have taken the position of an omniscient labeler, so all scenaria are labeled as they actually are, although the system may not be able to correctly label the scenario until many frames in the future.

   The main labeling difficulty is one of timing - when does one situation or scenario change into another. We have assumed that differences in this will be the sort of natural variation assessed as described above.

   The labeling of the roles/situations/scenaria was problematic. It was often unclear how each of the labels was to be used. We attempted to maintain at least consistent labeling by coordinating and reviewing of labels by one person. Therefore, the symbolic labeling is based on a best-guess representation of the final activity model.

3. **What is a group?**

   We have attempted to define a group as a set of individuals that are reacting to each other. This means that individuals may pass each other, *e.g.* one behind the other, without interacting and thus not forming a group. The human labelers can usually make this judgment, but it is less likely that an automatic labeler will be able to distinguish all instances of interaction. Thus, there is probably going to be a lot of false alarms on group box detection (*i.e.* individuals who are really not interacting, but just passing closely).

   Similarly, we grouped individuals that were interacting independently of the distance between the individuals, starting from the frame in which they first seemed to react to each other. For example, if two people wave while still quite distant and then turn to approach each other, the group box and labeling starts in the frame where the two noticed each other and initiated the waving.

4. **Multiple** *versus* **unique labels**

   Should an individual (or a group) have more than one role label, and participate in more than one situation and scenario at the same time? In labeling, we have decided only single classifications apply in each frame.

2. Semantic labeling

The modeled scenaria, their constituent situations, the participant roles allowed in each situation and the movement description for each role are summarized here.

The models are currently expressed as finite state automata, with the states as individual situations.

2.1 Plaza Observation Setting

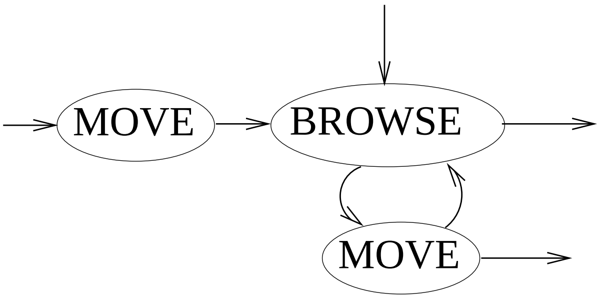
The different contexts that can give rise to scenaria are: Browse, Idleness, Drop-Dead, Walk, Fight, Meet, Leave-Object.

Solid ovals are individual situations, dashed ovals are group situations. Vertical bars are when two situations need to end at the same time.

For each scenario, there is a set of situations. Each situation (*e.g.* “Browse”) has listed the allowable Roles (*e.g.* “Browser”) and allowable Movements (*e.g.* “Inactive”): `BROWSE:Browser/{Inactive}`.

2.1.1 Browse Context

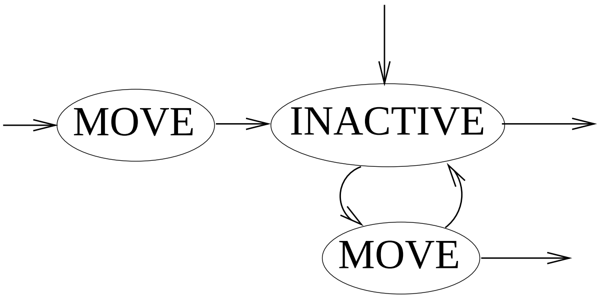
Actually looking at some information display:



MOVE: {Walker,Browser}/{Walking}
BROWSE: Browser/{Active,Inactive}

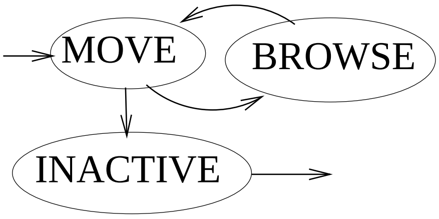
2.1.2 Idleness Context

Just standing around:



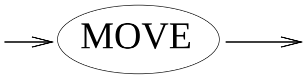
MOVE: Walker/{Walking}
INACTIVE: Walker/{Active,Inactive}

2.1.3 Drop Dead Context



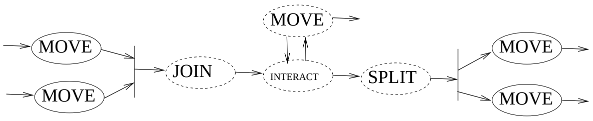
MOVE: {Walker,Browser}/{Walking}
INACTIVE: Walker/{Inactive}
BROWSE: Browser/{Active,Inactive}

2.1.4 Walk Context



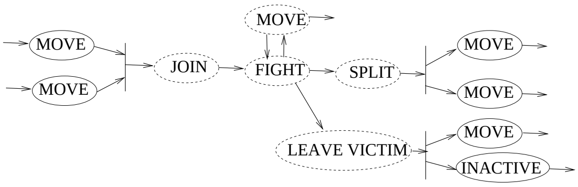
MOVE: Walker/{Walking}

2.1.5 Meet



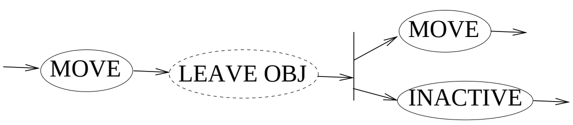
MOVE (individual): Walker/{Walking}
MOVE (group): Walkers/{Movement}
JOIN: Meeters/{Movement}
INTERACT: Meeters/{Active,Inactive}
SPLIT: Meeters/{Movement}

2.1.6 Fight



MOVE (individual): Walker/{Walking,Running}
MOVE (group): Walkers/{Movement}
JOIN: Fighters/{Movement}
FIGHT: Fighters/{Active,Movement}
SPLIT: Fighters/{Movement}
LEAVE VICTIM: Fighters/{Active,Movement}
INACTIVE: Left Victim/{Active,Inactive}

2.1.7 Leave-Object



MOVE (individual): Walker/{Walking}
INACTIVE: Left Object/{Inactive}
LEAVE OBJ: Walkers/{Inactive}

3. Shop observation scenario datasets

The web site given above will also eventually contain about 50 additional ground-truth labeled video sequences observing scenaria that occur in a shopping center, containing about 60000 labeled frames. This is expected to be complete in the summer of 2004.

Acknowledgements

Funding for this project was given under the EC's Information Society Technology's programme project IST 2001 37540. We'd like to thank many people for help with the ground-truth labeling: Toby Breckon, Helmut Cantzler, Ignasi Cos Aguilera, Jose Manuel Vazquez Diosdado, Dina Kronhaus, Gregor Miller and Donald Nairn.

A. Ground truth label grammar

The grammar and meaning of the ground truth files is as follows:

```
% the whole file
FILE -> FRAMELIST

% a list of frame descriptions
FRAMELIST -> FRAME
FRAMELIST -> FRAMELIST FRAME

% a frame description
FRAME -> frame NAME FID ibl IBLIST eibl gbl
    GBLIST egbl eframe

% a video sequence name
NAME -> character string with no blanks

% the frame number
FID -> an integer

% a list of individual boxes
IBLIST ->
IBLIST -> IBLIST IB

% a list of group boxes
GBLIST ->
GBLIST -> GBLIST GB

% an individual box description
IB -> ib IID IC IR IW IH IO IASL IFLAGL eib

  % individual box ID
  IID -> an integer

  % IR, IC - row and column of center of
  % individual box
  IC -> an integer
  IR -> an integer

  % IH, IW - height and width of individual
  % box
  IW -> an integer
  IH -> an integer

  % IO - main axis orientation [0..179]
  % degrees
  IO -> an integer

  % IASL - state flag list
  IASL ->
  IASL -> IASL IAS
  IAS ->
       ap            % appear
     | di            % disappear
     | o             % occluded
     | in            % inactive
     | ac            % active
     | wk            % walking
     | r             % running

  % IFLAGL - scenario flag list
  PROB -> a floating point probability
          in [0.0...1.0]
  IFLAGL ->
  IFLAGL -> IFLAGL IFLAG
  IFLAG ->
        f   PROB     % fighter role
     |  br  PROB     % browser role
     |  lv  PROB     % left victim role
     |  lg  PROB     % leaving group role
     |  wr  PROB     % walker role
     |  lo  PROB     % left object role
     |  m   PROB     % moving situation
     |  is  PROB     % insactive situation
     |  bsi PROB     % browsing situation
     |  bsc PROB     % browsing scenario
     |  im  PROB     % immobile scenario
     |  wg  PROB     % walking scenario
     |  dd  PROB     % drop down scenario
     |  pi  PROB     % immobile event

% a group box description
GB -> gb GID GC GR GW GH GO gibl GMEML egibl
      GASL GFLAGL egb

  % group box ID
```

```
GID -> an integer

% GR, GC - row and column of center of
% group box
GC -> an integer
GR -> an integer

% GH, GW - height and width of group box
GW -> an integer
GH -> an integer

% GO - main axis orientation [0..179]
% degrees
GO -> an integer

% GMEML - List of group members
GMEML ->
GMEML -> GMEML IID

% GASL - group state flag list
GASL ->
GASL -> GASL GAS
GAS ->
     ap            % appear
   | d             % disappear
   | i             % inactive
   | ac            % active
   | mo            % movement

% GFLAGL - scenario flag list
PROB -> a floating point probability
        in [0.0...1.0]
GFLAGL ->
GFLAGL -> GFLAGL GFLAG
GFLAG ->
      f   PROB     % fighters role
   |  me  PROB     % meeters role
   |  w   PROB     % walkers role
   |  gf  PROB     % fighting situation
   |  gmo PROB     % moving situation
   |  gme PROB     % meeting situation
   |  s   PROB     % split up situation
   |  gi  PROB     % inactive situation
   |  glv PROB     % leaving victim situation
   |  glo PROB     % leaving object situation
   |  fsc PROB     % fighting scenario
   |  mes PROB     % meeting scenario
   |  ls  PROB     % leaving object scenario
   |  fst PROB     % fight start event
   |  fe  PROB     % fight end event
   |  fv  PROB     % left victim event
```