The Understanding of Spatial-Temporal Behaviors

Yu-Jin Zhang
Tsinghua University, China

INTRODUCTION

The human visual process can be seen as a complex process ranging from feeling (feel the projection of 3-D world on the resulting 2-D image) to perception (capturing the content and meaning of 3-D world with 2-D images) (Kong, 2002). The ultimate goal of vision, from the narrow sense, is to make a meaningful interpretation of the world and the description of the scene; while broadly speaking, it includes also designing a behavior plan based on these explanations and descriptions, and in accordance with the surroundings and wishes of viewers. Computer vision is to realize the task to make computer refers to the human visual function, hoping to make a meaningful judgment on realistic goals and scenarios based on perceived images (Shapiro, 2001). This is just also the goal of understanding.

An important task in computer vision and image understanding is to analyze the scene through image operation on the image of scene in order to guide the action. To do this, one needs to locate the objects in the scene, and to determine how they change its position, attitude, speed and relationships in the space over time. In short, it is to grasp the action in time and space, to determine the purpose of the operation, and thus to understand the semantics of the information they passed. This is refereed as the understanding of spatial-temporal behaviors.

Behavior-based automatic target image/video understanding is a very challenging research issue. It includes the access of objective information (image/video acquisition sequence), the process of relevant visual information, the analysis of contents (expression and description), as well as the interpretation on the basis of the image/video information, in order to achieve learning and recognition of behavior.

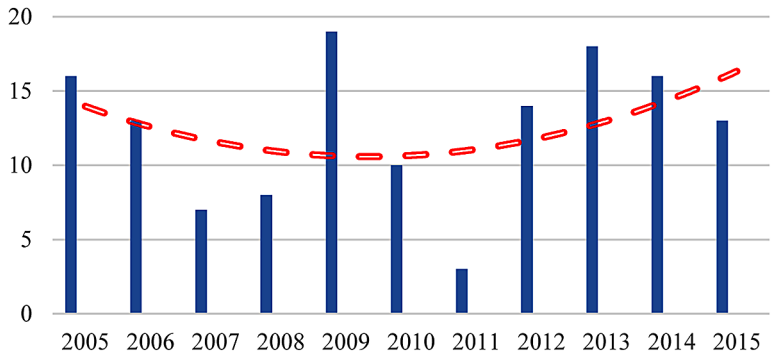
BACKGROUND

Research forces and results around such a topic are just appeared in recent years, some statistics can be seen from the survey on image engineering (Zhang, 2015c). The annual survey series of the yearly bibliographies on image engineering has started in 1995 and has been carried out for 21 years (Zhang, 2016). When the series enters its second decade (for the literature statistics of 2005), with the appearance of some new hot spots in the image engineering research and application, a new subcategories (C5): spatial-temporal technology (including 3-D motion analysis, gesture and posture detection, object tracking, behavior judgment and understanding) has been added into the image understanding category (C) (Zhang, 2006). The emphasis here is the comprehensive utilization of a variety of information possessed by the image/video in order to make the according interpretation for the dynamics of scene and objects inside.

In the past eleven years, the number of publications belong to the subcategory C5 in the annual survey series has attend a total of 153. There are five subcategories in category C, and the total number of publications belong to category C in these eleven years is 1352, so the subcategory C5

DOI: 10.4018/978-1-5225-2255-3.ch115

Figure 1. Some statistics of the numbers of publications for spatial-temporal technology



is still a small subcategory. Their distributions in each years are shown in the bars in Figure 1, in which a 3-order polynomial curve fitting to the number of publications of each year is drawn to show the change trends. Overall, this is still a relatively new field of research, so its development is not too stable, yet.

MAIN FOCUS OF THE ARTICLE

The definition, development, and stratification of spatial-temporal technology are first provided. Then, from low-level to high-level, the detection of points of interest, the forming of dynamic trajectory and activity path, the example techniques for action classification and recognition, as well as modeling for action and activity, are introduced consecutively. Several further research directions are discussed before some final concluding remarks are delivered.

Spatial-Temporal Technology

Spatial-temporal technology is oriented to the understanding of spatial and temporal behavior. Currently, the main target of the research on spatial-temporal technology is a moving person or object, and the object (particularly a human) changes. According to the expression level of abstraction and description, it can be divided into multiple levels from bottom to top.

1. **Action Primitives:** Atomic units to build a significant action, generally corresponds to a short period of motion information in the scene.
2. **Action:** A series of acts composed of a number of action primitives, a collection of ordered and meaningful models, generally represents a simple execution often carried out by one person, and often lasts only seconds in duration. The results of human actions often lead to changes in body posture and/or object position.
3. **Activity:** A series of actions performed by the host/initiator (mainly emphasize logic combinations). It is a relatively large-scale sports events, generally depends on the environment and people in interaction (Aggarwal, 2011). It is often an occurrence or a consequence of complex operations performed by more than one person, and often last for a relative long period of time.
4. **Event:** It refers to certain activities occurred in specific periods and specific spatial location. Typically one of the actions performed by a plurality of host / initiator (group activity). Detection of specific events often associated with abnormal activities, an overview on this can be seen from (Popoola, 2012).
5. **Behavior:** Subject/initiator mainly refers to a human or animal, emphasizing the body/originator dominated by ideological change action in a particular environment/context, the ongoing activities and describe events.