Object Motion Analysis and Interpretation in Video

Dan Song, Miyoung Cho Pankoo Kim
Dept. of Computer Science
Chosun University, Gwangju 501-759 Korea
{songdan, irune80, pkkim}@mina.chosun.ac.kr

Abstract

With the more sophisticated abilities development of video, object motion analysis and interpretation has become the fundamental task for the computer vision understanding. For that understanding, firstly, we seek a sum of absolute difference algorithm to apply to the motion detection, which was based on the scene. Then we will focus on the moving objects representation in the scene using spatio-temporal relations. The video can be explained comprehensively from the both aspects: moving objects relations and video events intervals.

1. Introduction

The objective of the computer vision is the interpretation of images what can we deduce about the world given one or more images of it? This understanding can be either at a high level-recognizing a person, or identifying an object-or at a low level-interpreting a picture, estimating motion. The input is image data, which can be either a single image, or a collection of images over time. From that, for low-level vision problems, we want to estimate an underlying scene, which could be 3-dimensional shape, or high resolution detail. We will focus on low-level scene representations like these that are mapped over space.

In order to analysis the video object motions in an event At the beginning, we prototype a complete video application in simulink using the frame-based, multirate, matrix-oriented signal processing capabilities in the digital signal processor (DSP) blockset[1]. We have tried to approach new video processing tasks in simulink for the motion detection. We first examine data compression, and then describe how to prototype and simulate a motion-detection algorithm for a video surveillance application with simulink and the DSP blockset.

The other important part in this paper is how to describe and interpret the moving objects. participanting in the video event based on the scene. Our approach is related to the work from Raines an d his colleagues who describe a new approach to au tomate assistants to aid humans in understanding te am behaviors for the simulation league[2]. This app roach is designed for the analysis of games, off-lin e after playing, gain new experiences for the next-games. Frank and colleagues presented a real time approach which is based on statistical methods[3]. A team will be evaluated statistically but there is no recognition of team strategies. Object motion takes place in space and time. Therefore, it is useful to describe the behavior of moving objects in terms of spatio-temporal relations. This leads to a domain independent description. Once the spatio-temporal relations between the objects are described one can interpret moving object in the scenario.

2. Simulink Video Processing for Object Motion Detection

Video surveillance systems produce large amounts of data to record, archive, and review, data compression is a vital part of next-generation surveillance systems. For example, even a modest 8-bit/pixel monochrome digital surveillance system with common intermediate format (CIF) resolution (352x240 pixels) running at 30 frames/second will send approximately 2.4 MB/second to a storage device.

There are two primary ways to decrease the space required to store the surveillance video data: store only those video frames that have certain characteristics, such as significant motion; or store compressed representations of individual video frames. In this paper, we use a simple motion-detection algorithm to automatically select individual video frames for storage. This will help us reduce our example's recording rate. One of the simpler algorithms for estimating motion is to compute the difference between successive video frames.

The greater the number of differences that we find, the more motion is likely to have occurred within the scene. We designed the block diagram in Figure 1 to compute the Sum of the Absolute value of Differences (SAD) between successive video frames, returning a useful estimate of motion. In Figure 1, output port #2 (SAD) produces the motion estimate.

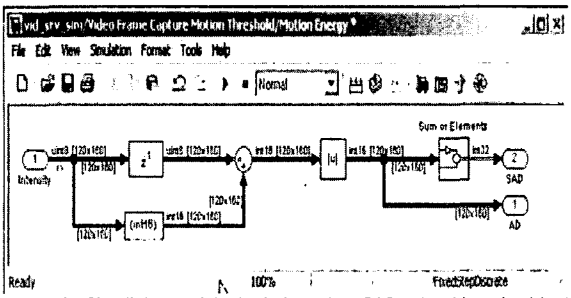

Fig.1 A Simulink model depicting the SAD algorithm in block diagram form

The video source is put into port #1, and the operations use integer data types for memory efficiency. The auto-scaling and overflow-reporting features and fixed-point parameters of simulink help with investigating fixed-point effects, such as detecting dynamic range overflow conditions or assessing the impact of rounding modes on intermediate computations.

3. Video Object Motion Description Using Spatio-Temporal Relations

3.1 Spatio-Temporal properties and relationships of video object

Scenarios description with moving objects is to identify the concepts a human observer uses to describe the movement of the objects. In this paper we split the concept into the video object motion interval (VOMI), and spatial relation alternation (SRA), which have to be generated to step by step at run-time. And the temporal relations between these time intervals are described. This leads to a definition of the concept in terms of spatio-temporal relations and make it possible to identify the concept within a scenario[4].

Elementary concepts describing simple motion events are domain independent and build a basis for the construction of more complex events. Complex events are often domain specific and are described as a combination of simple events. So, each VOMI refers to exactly one object O and has exactly one start moment i_s, end moment i_e, motion direction α and motion speed v: $i_{vom} = [O(\alpha, v)]_{i_s}^{i_e}$. Each SRA refers to exactly one pair of objects (O_1,O_2) and has exactly one start moment a_s, end moment a_e, location direction l and displacement d: $a_{sr} = [O_1 < l, d > O_2]_{a_s}^{a_e}$. The relationship between the time intervals - VOMI and SRA- are described using seven temporal relations before, meets, overlap, contains, starts, finishes and equal and their inverse described in [Allen, 1981][5] as shown in Figure2.

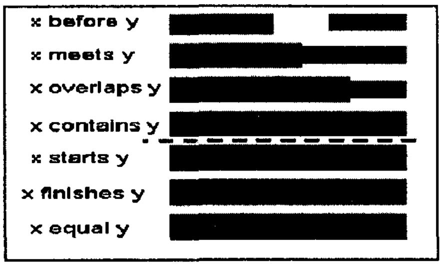

Fig.2. Visual presentation of relationships in a 2-dimensional temporal space

3.2 Video Object Motion Description

Any scenarios of moving objects can be described by use of VOMI and SRA which are temporally related. Here, we consider the scene(Figure.3)-"**The player is approaching the ball**"-obtained through the motion detection algorithm in the video of "**Penalty Goal**" as our example to explain the event.

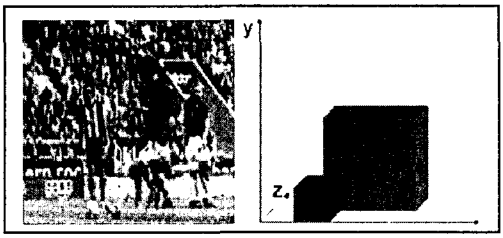

Fig.3. Capture 14: The player is approaching the ball

From the Figure 3 we know that the event of " Penalty Goal " is constituted by a series of continuous scenes just like "Capture 14", which was the motion detected from the source video. The right graph showed us the spatial relationships between the "Player"-A and the "Football"-B in the space. However, if we want to describe an event, not only should we explain the relative position relationships in the space, but also expose the time intervals[6]. So, we will use the seven kinds of temporal alternations to interpret the spatial relationships and the time intervals. This motion can be explained from three directions: X-direction, Y-direction and Z-direction. The interactions and relations will be shown is the Figure4.

x: A overlapped by B
y: A contains B
z: A met by B

$A\langle ov - by, con, met - by \rangle B$

Fig.4.Temporal relations describe the object motion

From the Figure4, this "approaching" motion interacting between the football and the player in this specific time can be described by the combinations of these three different kinds of temporal relationships.

Next, we will take advantage of our approach to interpret how objects are spatially and temporally related to others and track this over the time in the scene of Capture 14-"The player is approaching the ball". We define the "player" as the O_1 and "football" as the O_2. Two objects including in the simple event that are spatial related and may move within duration of the spatial relation, i.e. three overlapping time intervals are involved:

$a_{sr} = [O_1 < l_1, d_1 > O_2]$, $i_{vom1} = [O_1 < \alpha_1, v_1 >]$ $i_{vom2} = [O_2 < \alpha_2, v_2 >]$ ---- (1)

To distinguish and recognize the different simple events belonging to one group the attributes motion direction, speed, direction of spatial location and distance have to fulfill certain constraints, e.g.:

- Two objects meet each other: d_1 = meets.
- O_1 is approaching O_2 ($app(O_1,O_2)$): $v_1 > 0$ ^ $\alpha_1 = l_1$ ^ $v_2 = 0$.
- O_2 is departing from O_1 ($dep(O_1,O_2)$): $v_2 > 0$ ^ opposite(α_2, l_2) ^ $v_1 = 0$.

In the same manner it is possible to define constraints for objects moving in parallel or following each other. In addition to this there are several further groups of simple events that are not mentioned in this paper.

4. Conclusion

In this paper, our final goal is to analyze and interpret the object motions in the live enviroment. So, first we use the technology of simulink video processing for motion detection. Especially, sum of absolute difference algorithm has played an important role in our experiment. We also have put forward a common sumulink model for the motion estimation.

Furthermore, our challenging task is to describe the video object motion within real-time environment. We present our approach to track the objects and therefore the spatio-temporal relations between them on line so that can interpret this situation. We showed that simple events such as meets, departing, approaching, equals can be detected and combined to more sophisticated events. Also, we are able to detect relations between two objects but sometimes more than two objects are involved in a situation, This will be on the list for our future work.

Reference

[1] Juan W Dixon and Ivan A Leal, 2002. "Current Control Strategy for Brushless DC Motor Based on a Common DC Signal". IEEE Transactions on Power Electronics, Vol 17(2).

[2] Raines, T., Tambe, M., and Marsella, S. (2000). Automated assistants to aid humans in understanding team behaviors. In Fourth International Conference on Autonomous Agents (Agents 2000), Barcelona, Spain.

[3] Frank, I., Tanaka-Ishi, K., Arai, K., and Matsubara, H. (2000). The statistics proxy server. In Balch, T., Stone, P. and Kraetschmar, G., editors, 4th International Workshop on RoboCup, pages 199-204, Melbourne, Australia. Carnegie

[4]Fernando, T. (2003). Finite-state descriptions for temporal semantics. In H. Bunt, I. van der Sluis and R. Morante (Eds.), Proceedings of The Fifth International Workshop on Computational Semantics (IWCS-5), 122-136, Tilburg, The Netherlands, January. Mellum University Press.

[5] Allen, J. F. (1981). An interval-based representation of temporal knowledge.In Hayes, P. J., editor, IJCAI, pages 221-226, Los Altos, CA.

[6] Visser, U., Dr¨ucker, C., H¨ubner, S., Schmidt, E., and Weland, H.-G. (2001). Recognizing formations in opponent teams. In RoboCup-00, Robot Soccer World Cup IV, Lecture Notes in Computer Science, Melbourne, Australia. Springer-Verlag. to appear.