Trajectory Mapping for Landmine Detection Training

Yang Cai

Human-Computer Interaction Institute,
School of Computer Science,
Carnegie Mellon University,
5000 Forbes Avenue, Pittsburgh, PA 15213, USA
`ycai@cmu.edu`

Abstract. A head-mounted camera is a useful tool for studying the usability of mobile devices in the field. In this paper, a computerized visualization method is presented. It includes the target trajectory mapped with the deformable template-based tracking algorithm and landmarks-based relative object registration. A landmine detection training video is used for the case study. The results show that this approach has advantages over optical flow and overhead camera methods.

1 Introduction

Human field performance has been studied for decades, from golfing to landmine detection. It has become a renaissance area because of: 1) emerging mobile computers for field applications, such as Ground Penetration Radar (GPR) for landmine detection and handheld training computers for Navy personnel, etc., 2) emerging remote control through telepresence, such as robotic rescue systems and capsule medical cameras, 3) traditional manned field missions with new situations, such as landmine detection for peace-keeping and vehicle driving studies, etc. In light of this, human field performance study is an "old field" that is redeemed with new technologies.

Video cameras have been widely used in human performance studies, such as surveillance camera, infrared camera, high-speed camera, microwave imaging camera, etc. It is common to keep vision systems static while tracking human subjects' movement. In these cases, human tracking is relatively easy during the video post-processing phase. For example, we can use the background subtracting method to separate the human subjects and the static background. Also, it is easy to measure the distance or track the motion speed. However, in many situations, the static camera-based approach is rather expensive or difficult to use in the field. For example, it is very hard to use a single static camera to track human activities in an obscured scene or multiple rooms. In addition, it is hard to track the human operation in a very large open field without an overhead camera or multiple cameras. If we use multiple cam-

eras, then we have to add image fusion, object registration and synchronization functions. In these cases, a head-mounted video camera seems a reasonable choice.

Head-mounted video cameras have been used in special research projects for example, "augmented reality" by video overlaying on head-mounted displays [1,2,3,4,5,6], lip movement tracking [7] and eye movement tracking [8]. The "augmented reality" registers computer generated graphics to a video image. In this paper, we attempt to do the opposite: to register an object from the video image to computer generated graphics. The current head-mounted camera-based lip and eye movement tracking systems only observe a single component on human face. In this paper, the author focuses on a broader problem: "how to visually evaluate human field performance with a head-mounted video camera?" The goal of this study is to develop a computerized object tracking and mapping system that can automatically register the moving target to a trajectory map. This study uses military landmine detection video as a case study and uses computer vision algorithms to map the original video data to a dynamic tracking graph. It is expected that the method can be applied to other fields, such as behavior measurement for elderly in nursing homes, user performance modeling for airplane inspection, etc.

2   Trajectory Map

A trajectory map is a 2D or a 3D space that is registered with a dynamic trace of a target. It is a visual model of human field performance dynamics, for example, search patterns, pace, and sweeping patterns, etc. Fig.1 shows a comparison of a trace of a sweeping metal detector from an expert and a novice. From the map we discovered that the trace of the expert is uniform and thorough. However, the trace of the novice is uneven that contains missing spots.
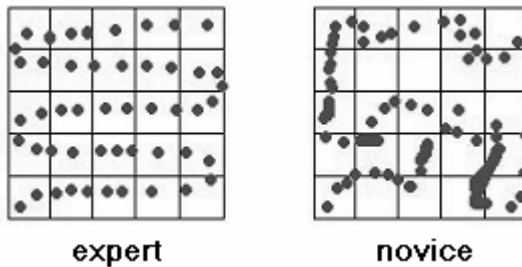

expert   novice

Fig. 1. *"Sweeping Patterns" of a metal detector head. Note the expert's pattern is uniform and thorough. The novice's pattern is uneven and random*

Visualization helps the analysts to understand the dynamics of the human behavior in the trajectory. The methods include: 1) time-stamping the trajectory points with "temperature colors", which maps a duration time with a color, etc., 2) plotting the effec-

tive halo envelope for the trajectory, which reveals the overlapping patterns. In many cases, adding verbal protocols which are aligned to the trajectory points would help analysts to understand the subject's motivation, cognition and decision making process.

The challenge for processing the data from a head-mounted video is how to register the trajectory map. A head-mounted camera has four degrees of freedom (DOF): pitch, yaw, tilt and zoom, which make the registration rather cumbersome. To make efficient target tracking and registration, landmarks are recommended in the head-mounted video tracking. Fig 2 shows an example of the usage of the measurement tapes as landmarks.
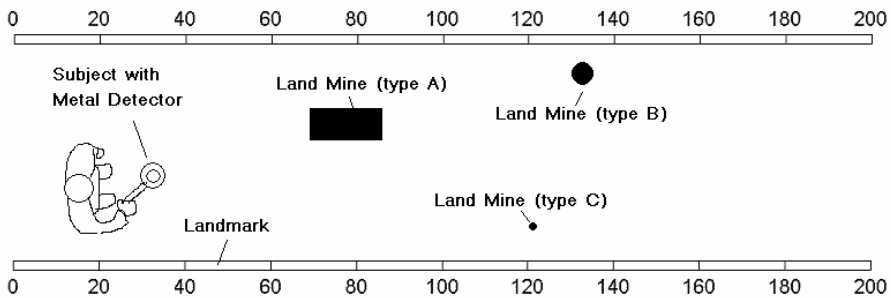

Fig. 2. *Landmine detection training field*

3   Target Tracking

The moving target normally has a defined shape, e.g. circle, etc. In this study, an active contour model, "Snake" [9] is used for tracking. It is a general algorithm for matching a deformable model to an image by means of energy minimization. The energy function is a weighted combination of internal and external forces. The snake is defined parametrically as $\mathbf{v}(s) = [x(s), y(s)]$, where $x(s)$, $y(s)$ are x, y co-ordinates along the contour and $s \in [0,1]$. The energy functional to be minimized as

$$E_{snake} = \int_0^1 \{[E_{int}(\mathbf{v}(s)] + [E_{image}(\mathbf{v}(s)] + [E_{con}(\mathbf{v}(s)]\} \, ds$$

where E_{int} represents the internal energy of the spline due to bending, E_{image} denotes image forces, and E_{con} external constraint forces. In this case, $\mathbf{v}(s)$ is approximated as a spline to ensure desirable properties of continuity.

Fig.3 shows a result from the algorithm for tracking a metal detector head from the head-mounted video. After nearly ten iterations for each frame, the deformable "snake" successfully located the metal detector head in the video.

Fig. 3. *Metal detector head tracking result. The square dots indicate the tracked target*

4   Target Registration

There are many ways to register a target to a two-dimensional map. Because a head-mounted camera has at least four degrees of freedom (DOF), the images in the video are geometrically distorted. A classical way to correct the image is linear or non-linear transformation. With a few pairs of "control points", a warping function can be used to transform the distorted raw images from pixels in the source video to the destination defined by a transformation function. [10] We call the transformation-based methods "*absolute registration*" methods since they generate "absolute" coordinates of the tracked object in the map.

However, in reality, there are only very few video frames containing required control points for the linear transformation. In many cases, the video only shows partial landmarks. What do we do to fill in these gaps? In addition, it is not necessary to transfer all pixels from one to another. For target tracking purposes, we only need the relative coordinators and distances referenced to the landmark and the plane of the landmarks. In light of the shortcomings of the transformation, this study focuses on the "*relative registration.*"

Relative registration is a non-metric measurement method in which a target is not only just an object but also a reference itself. It is an approximate way for a quick measurement of the object's size and the distance between things. For example, artists often use "number of heads" to measure human figure's height and use "number of eyes" to measure the width of face. This is based on observations of our daily life. Our perception systems have "internal yardsticks" for qualitative measurement. Our eyes do not make absolute measures of characteristic of the subject, but instead detect these characteristics only in a relative way. We do not see the true color of a thing, but rather an apparent color, which is our sensation of how a color is different from the colors surrounding it. As Van Gogh said, "There is no yellow if there is no blue if there is no red." The same principle applies to our perception of edges, patterns, and shapes. We may focus our attention solely on a part of the subject or notice this part

peripherally as we scan our eyes around the whole visual field. More evidence has been found from artists' painting textbooks and the landmine detection tapes.

The relative registration procedure in this study is based on a few assumptions. For example, we assume that the metal-detector head is near parallel to the ground and the subject looks at the near field ground, etc. The registration heuristics is as followings:

- Determine a template (box) of the object (metal detector head) in the image.

- Locate a feature point on the landmark (e.g. the numerical mark)

- Find the reference lines that are either perpendicular or parallel to the land-mark on the plane.

- Measure the distances (perpendicular and parallel directions) from the object (metal-detector head) to the landmark with the template. For example, in Fig. 4 on the right, the relative coordinate to the mark "100" for the metal-detector head object is (0.8,0) in terms of "number of templates"
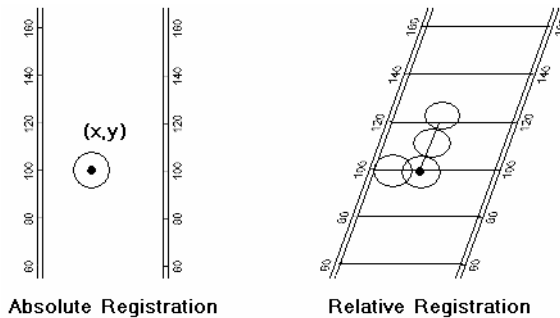

Absolute Registration   Relative Registration

Fig. 4. *Illustration of "Relative Registration" versus "Absolute Registration." In relative registration, the template of the target is used as a measurement reference.*

5   Results

A preliminary experiment has been conducted based on a 60 minute test video from the field. It was stored on MiniDV and converted to AVI files at 30 frame per second with the frame size of 177 x 172. Before the video analysis, we removed about 20% of "irrelevant" clips, e.g. vomiting and resting. Those irrelevant clips are easy for humans to understand but difficult for computational processing. Fig.5 shows a result of the trajectory mapping from a video clip, where the dots are reference points with inter-vals of 10 frames and the trace of the metal detector head indicates a normal sweeping pattern, which is uniform and thorough. Fig.6 and Fig. 7 show examples of the traces of a metal detector head while making decisions to determine the location and type of the landmine. The closer dots, the more decision time that the subject spent. Also, from the trajectory maps, we found a few decision making "styles." For example, the pattern in Fig.5 shows a circular search style. The pattern in Fig.6 shows a cross-shape search style.
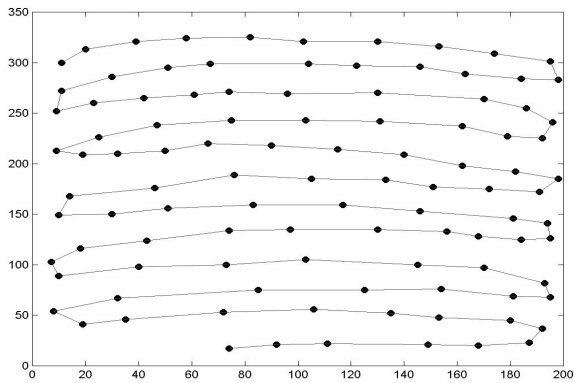
Fig. 5. *An output of the "sweeping" patterns. Dots are reference points with an interval of 10 frames and the trace of the metal detector head indicates a normal sweeping pattern, which is uniform and thorough.*
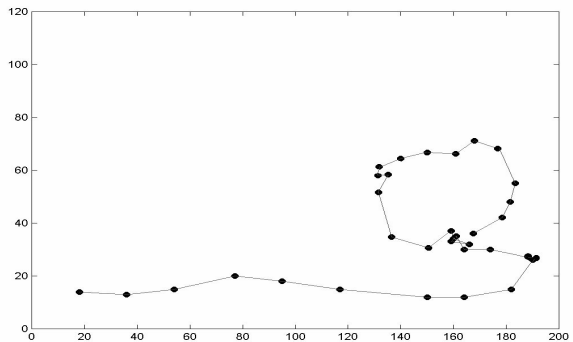

Fig. 6. *A circular shape searching style. The trace of the metal detector shows how the subject makes decisions to determine the location and type of landmines.*
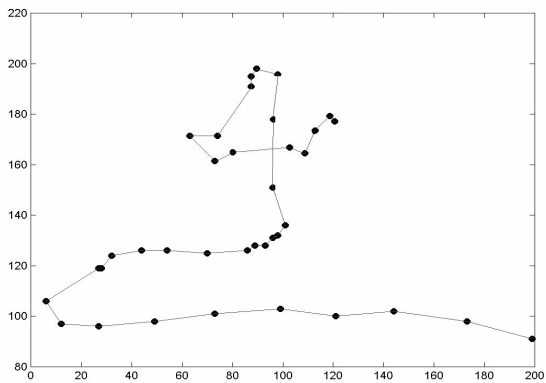

Fig. 7. *A cross-shape searching style. The trace of the metal detector shows how the subject makes decisions to determine the location and type of landmines.*

Table 1. *Performance summary*

Items	Performance
Registration accuracy	¼ to ½ of metal detector head diameter
Trajectory mapping speed	Near real-time
Object tracking speed	3 times real-time on PC (700 MHz,256MB RAM)
Landmark visibility	57% single lane, 19% both lanes, 24% none
Manual reinitiating	35% (e.g. subject looks at sky, drinks water, etc.)

Table 2. *Comparison between head-mounted camera and overhead camera*

Items	Head-Mounted Camera	Overhead Camera
Protocol sound tracks	yes	Wireless needed
Camera tower	no	yes
Object registration	Relative registration	Absolute registration
Object tracking	Software-based	Light-bulb-based
Distance distortion	less	yes

Table 3. *Comparison between the Landmark-based and Optical Flow-based methods*

Items	Landmark-based	Optical Flow-based
"drafting"	no	yes
generalization	no	yes
computation	simple	intensive

6   Discussions

A preliminary experiment has been also conducted to apply an Optical Flow algorithm to estimate the moving directions. Unfortunately, because there is a moving target in addition to the moving background, it is rather hard to separate the target from the background.

At this stage, manual initiation is used in tracking. Since the manual initiation is just to put a circular shape near the center of the target, and the manual and automation ratio is significant, it is tolerable to use the manual initiation.

The main problem for tracking based on the head-mounted camera is the "drifting problem". The trajectory would accumulate the error by time and eventually drift away from the correct course. To avoid the drifting problem, we have to use frequent

landmark checking, introduce additional sensors, such as magnetic sensors, etc. to calibrate the ground truth at a certain duration. Besides, dual-camera registration, e.g. one for head-mounted and another for overhead, is desirable to minimize drifting.

Tracking a target with the current algorithm is time-consuming. The video processing time over the real-time is 3:1 on a PC with 700MHz, 256MB RAM. To improve this, development of faster tracking algorithms is underway.

At this stage, the optimal tracking accuracy is ¼ to ½ of the size of the metal detector head template. A 2D trajectory map may also lose the 3D information. For example, landmine detection experts sometimes turn the metal detector head at 45 degree to identify the halo envelop of the mine or just test the head. It is hard to observe this on a 2D map.

7   Conclusions

In this paper, a visual trajectory model for field performance evaluation is presented. The author addresses the approach for a trajectory map, target tracking and the "relative object registration" method. A landmine detection training video is used as a case study.

Although head-mounted video cameras have been used for field performance observation for long time, very little has been done in trajectory mapping. It is concluded that the "snake" algorithm can be used for deformable target tracking for the head-mounted video camera. However, it is rather computation intensive. The *relative registration* method is a novel approach to map the target from incomplete larndmarks in video scenes. It provides approximate and fast qualitative measurement. Compared to the overhead camera and optical flow methods, this approach is inexpensive and flexible.

It is also suggested that computerized head-mounted video analysis has its great potential in studying human field performance. To improve the accuracy of the image registration, physical trackers, such as magnetic sensors, are recommended in addition to the landmarks in the video. In addition, since video analysis involves both human and computational collaborative efforts, a well-designed human-computer interface would significantly increase the productivity.

References

1.   Chai, L., W.A. Hoff, and T. Vincent, "3-D motion and structure estimation using inertial sensors and computer vision for augmented reality," *Presence: Teleoperators and Virtual Environments*, 2000.

2.   Chai, L. and K. Nguyen, W. Hoff, and T. Vincent, "An adaptive estimator for registration in augmented reality," *Proc. of 2nd IEEE/ACM Int'l Workshop on Augmented Reality*, San Francisco, Oct. 20–21, 1999.

3.   Hirota, G. et al, "Hybrid tracking for augmented reality using both camera motion detection and landmark tracking", US Patent 6064749, May 16, 2000

4.   Hoff, W. A. and T. Vincent, "Analysis of Head Pose Accuracy in Augmented Reality," *IEEE Trans. Visualization and Computer Graphics*, Vol. 6., No. 4, 2000.

5.   Hoff, W. A. "Fusion of Data from Head-Mounted and Fixed Sensors," Proc. of First International Workshop on Augmented Reality, IEEE, San Francisco, California, November 1, 1998.

6.   Hoff, W. A., Lyon, T., and Nguyen, K., "Computer Vision-Based Registration Techniques for Augmented Reality," Proc. of Intelligent Robots and Computer Vision XV, Vol. 2904, in Intelligent Systems & Advanced Manufacturing, SPIE, Boston, Massachusetts, Nov. 19-21, pp. 538–548, 1996.

7.   Takaaki, K, et al. "Principal Components Based Lip Contour Extraction from Head-Mounted Camera and Cross-Subject Facial Animation", IPSJ SIGNotes Computer Graphics and cad Abstract No.100 - 014, 2000

8.   Sodhi, M., B. Reimer, JL. Cohen, E. Vastenburg, R. Kaars, S. Kirchenbaum. "On-Road Driver Eye Movement Tracking Using Head-Mounted Devices". Proceedings of the Eye Tracking Research and Applications Symposium, March 2002

9.   Trucco, E. and Verri, A. "Introductory Techniques for 3-D Computer Vision", Prentice Hall, 1998

10. Lillesand, T. et al, Remote Sensing and Image Interpretation, fourth edition, John Wiley & Sons, Inc. 2000