Automated Camera Selection and Control for Better Training Support

Adrian Ilie[1] and Greg Welch[2]

[1] The University of North Carolina at Chapel Hill
[2] The University of Central Florida

Abstract. Physical training ranges have been shown to be critical in helping trainees integrate previously-perfected skills. There is a growing need for streamlining the feedback participants receive after training. This need is being met by two related research efforts: approaches for automated camera selection and control, and computer vision-based approaches for automated extraction of relevant training feedback information.

We introduce a framework for augmenting the capabilities present in training ranges that aims to help in both domains. Its main component is ASCENT (Automated Selection and Control for ENhanced Training), an automated camera selection and control approach for operators that also helps provide better training feedback to trainees.

We have tested our camera control approach in simulated and laboratory settings, and are pursuing opportunities to deploy it at training ranges. In this paper we outline the elements of our framework and discuss its application for better training support.

1 Introduction

In recent years, physical training ranges have proven instrumental in providing trainees with a way to integrate skills perfected separately in an environment that is similar to the operational environment. Training feedback is provided in the form of After Action Reviews (AARs), which currently require a large number of highly-experienced instructors to accompany different segments of the unit throughout their training run.

Some training ranges have been equipped with large networks of hundreds of cameras, which can capture training exercises as they take place. Many cameras have pan-tilt-zoom (PTZ) capabilities, and are manually controlled by operators during the exercises. Other cameras are static, but operators still need to manually select which cameras to record, because only a limited number of recording devices are usually available. To alleviate these problems, automated approaches are being pursued to augment the operators' capabilities in controlling PTZ cameras and selecting which video streams to record.

The availability of cameras and operators has enabled instructors to provide a package containing multiple hours-long video segments manually selected by the operators. However, in order to pinpoint problem areas, the videos in the

D.D. Schmorrow and C.M. Fidopiastis (Eds.): AC/HCII 2013, LNAI 8027, pp. 50–59, 2013.

package need to be reviewed in their entirety. Computer vision algorithms can be employed to analyze the captured images and automatically extract information relevant for training feedback.

The framework introduced in this paper supports these efforts through ASCENT, an automated camera selection and control approach designed to support camera operators, while also helping provide better feedback to trainees by taking into account the requirements of the computer vision algorithms that process the captured images.

ASCENT consists of a *stochastic performance metric* and a *constrained optimization method*. The performance metric quantifies the uncertainty in the state of the targets. It can account for occlusions, accommodate requirements specific to the algorithms used to process the images, and incorporate other factors that can affect their results. The optimization method explores the space of camera configurations over time under constraints associated with the cameras, the predicted target trajectories, and the image processing algorithms. To achieve real-time performance, it combines a global assignment of cameras to targets that divides the problem into subproblems with a local optimization inside each subproblem. The global assignment uses a proximity-based heuristic to group targets and a greedy heuristic based on performance metric evaluations to assign cameras to each target group. It can also perform camera selection when needed. The local optimization is performed at the level of each group. It predicts the trajectories of all targets in the group and plans dynamic camera configurations over time to ensure optimal coverage up to a time horizon. While only some of the available cameras may be selected for recording, all captured images are available for algorithms that run in real-time, some of which can even provide feedback to ASCENT.

We have applied ASCENT to simulated and laboratory settings, and are pursuing opportunities to deploy it at training ranges that have already been outfitted with large camera networks. Our framework is well-positioned to help augment training capabilities. First, it augments camera operators' capabilities, allowing them to more effectively manage large camera networks. ASCENT automates camera selection and control decisions, allowing operators to either direct it to cover important events, or directly manage a smaller number of cameras. Additionally, ASCENT can be customized to produce images best-suited for the computer vision approaches that analyze and help extract relevant training feedback data. This has the potential to shorten AAR video packages down to automatically-selected segments that can be reviewed much faster.

The rest of the paper is organized as follows. In Section 2 we present some relevant research: a few performance metrics and camera control methods, as well as a few computer vision approaches that can be used to augment training. Section 3 presents our approach to camera selection and control: our performance metric and our camera selection and control method, as well as some experimental results. Section 4 briefly describes our framework and its potential contributions to better training support. We discuss some future work and conclude the paper in Section 5.

2 Previous Work

2.1 Performance Metrics

Many researchers have attempted to express the intricacies of factors such as placement, resolution, field of view, focus, etc. into metrics that could measure and predict camera performance. Below we list the performance metrics research closest to our work. The interested reader can find a comprehensive list of camera performance metrics in Chapter 2 of [10].

Allen [1] introduces steady-state uncertainty as a performance metric for optimizing the design of multi-sensor systems. In previous work [9] we illustrate the integration of several performance factors into this metric and envision applying it to 3D reconstruction using active cameras.

Denzler et al. [3] derive a performance metric based on conditional entropy to select the camera parameters that result in sensor data containing the most information for the next state estimation. In [4], Denzler et al. present a performance metric for selecting the optimal focal length in 3D object tracking. The determinant of the a posteriori state covariance matrix is used to measure the uncertainty derived from the expected conditional entropy given a particular action. Visibility is taken into account by considering whether observations can be made and using the resulting probabilities as weights. The authors of Deutsch et al. [6,5] improve the process by using sequential Kalman filters to deal with a variable number of cameras and occlusions, predicting several steps into the future and speeding up the computation. The ASCENT performance metric presented in Section 3.1 is similar to the metric by Denzler et al., but it uses a norm of the error covariance instead of entropy as the metric value, and employs a different aggregation method.

2.2 Camera Selection and Control Methods

Camera selection and control methods are typically encountered in surveillance applications. Many are centralized approaches, based on the adaptation of scheduling policies, algorithms and heuristics from other domains to camera control. Others are distributed: decisions are arrived at through contributions from collaborating or competing autonomous agents. We list a few example methods below. The interested reader is referred to Chapter 2 of [10] for a comprehensive list.

Qureshi and Terzopoulos [19] propose a virtual testbed for surveillance algorithms and use it to demonstrate two adapted scheduling policies: first come, first serve (FCFS) and earliest deadline first (EDF). In [18], they apply the same paradigm to a distributed surveillance system, in which cameras can organize into groups to accomplish tasks using local processing and inter-camera communication with neighbors in wireless range.

Naish et al. [17] propose applying principles from dispatching service vehicles to the problem of optimal sensing. They present a dynamic dispatching methodology that selects and maneuvers subsets of available sensors for optimal data

acquisition in real-time. The goal is to select the optimal sensor subset for data fusion by maneuvering some sensors in response to target motion while keeping other sensors available for future demands.

Lim et al. [13] propose solving the camera scheduling problem using dynamic programming and greedy heuristics. The goal of their approach is to capture images that satisfy task-specific requirements such as: visibility, movement direction, camera capabilities, and task-specific minimum resolution and duration.

Krahnstoever et al. [11] present a system for controlling 4 PTZ cameras to accomplish a biometric task. Scheduling is accomplished by computing camera plans: lists of targets to cover at each time step. Plans are evaluated using a probabilistic performance objective function to optimize the task success probability.

Broaddus et al. [2] present *ACTvision*, a system consisting of a network of PTZ cameras and GPS sensors covering a single connected area that aims to maintain visibility of designated targets. Cameras are tasked to follow specific targets based on a cost calculation that optimizes the task-camera assignment and performs hand-offs from camera to camera. The authors develop optimization strategies to either use the minimum number of cameras needed, or encourage multiple views of a target for 3D reconstruction.

Sommerlande and Reid [21] present a probabilistic approach to control multiple active cameras observing a scene. Similar to the approach in ASCENT, they cast control as an optimization problem, but their goal is to maximize the expected mutual information gain as a measure for the utility of each parameter setting and each goal. The approach allows balancing conflicting goals such as target detection and obtaining high resolution images of each target.

Matsuyama and Ukita [15] describe a distributed system for real-time multi-target tracking. The system is organized in three layers (inter-agency, agency and agent), with agents that dynamically interchange information with each other.

2.3 Computer Vision Approaches to Augment Training

There are many computer vision approaches that can process images, ranging from posture recognition from single images [22] to full 3D reconstruction from multiple images: multi-view dynamic scene modeling [7], space carving [12], 3D video [14] and image-based visual hulls [16]. However, most of these approaches have yet to be applied to large environments such as training ranges. Moreover, there are few approaches that can analyze the results of computer vision algorithms and extract relevant information that can help augment training. Sadagic et. al. [20] describe a concerted research effort in this direction. ASCENT provides ways to take into account the requirements of these approaches in order to capture images that are likely to produce the best possible result.

3 Automated Camera Selection and Control

We approach camera selection and control as an *optimization problem* over the space of possible *camera configurations* (combinations of camera settings) and

over time, under constraints derived from knowledge about the cameras, the predicted target trajectories and the computer vision algorithms the captured images are intended for. The objective function is a performance metric that evaluates dynamic, evolving camera configurations over time. In this section, we briefly describe the two components of ASCENT: its camera performance metric and its camera selection and control method. The interested reader is referred to [8] and Chapters 5 and 6 of [10] for a detailed presentation.

3.1 Camera Performance Metric

We define the performance of a camera configuration as its ability to resolve 3D features in the working volume, and measure it using the uncertainty in the state estimation process. We use state-space models [8] to describe target dynamics and measurement systems. Formally, at time step t, the system state is described by a *state vector* $\bar{x}_t \in \mathbb{R}^n$ which may include elements for position, orientation, velocity, etc. Given a point in the state space, a mathematical *motion model* can be used to predict how the target will move over a given time interval. Similarly, a *measurement model* can be used to predict what will be measured by each sensor. We measure the uncertainty in the state $\bar{x}_t$ using the a posteriori error covariance P_t^+, which we compute by applying the Kalman Filter equations to elements of the state-space models.

Our performance metric evaluates *plans*: temporal sequences of camera configurations up to a *planning horizon*. We compute the performance metric for each candidate plan by repeatedly stepping forward in time up to the planning horizon, while applying the Kalman Filter equations and changing relevant state-space model parameters at each time step. We use the motion models to predict target trajectories and generate predicted measurements, and we update the measurement models with the camera parameters corresponding to the configurations planned for each time step. We aggregate over space and time using weighted sums, with weights quantifying the relative importance of elements at various levels, such as points in a target surrogate model, targets, or time instants. Equation 1 illustrates the general formula for the metric computation using weighted sums.

$$\mathcal{M} = \sum_{r=1}^{N_t} u_r \left(\sum_{t=1}^{H} v_t \left(\sum_{p=1}^{N_r} w_p \left(\sqrt{Max\left(Diag_{pos}\left(P_{t,p}^+\right)\right)} \right) \right) \right) \tag{1}$$

N_t is the number of targets, N_r is the number of points in the surrogate model of target r, H is the planning horizon. u_r, v_t and w_p are relative weights for each target r, time step t, and model point p, respectively. $P_{t,p}^+$ is the a posteriori covariance for model point p at time t. To convert the error covariance into a single number, we use the square root of the maximum value on the diagonal of the portion of the error covariance matrix $P_{t,p}^+$ corresponding to the position part of the state.

3.2 Camera Selection and Control Method

We define optimization in active camera selection and control as the exploration of the space of possible solutions in search for the best solution as evaluated by the performance metric. Our optimization process first predicts the target trajectories, then uses them to construct and evaluate a number of candidate plans for each camera. A plan consists of a number of *planning steps*. A step consists of a *transition* (during which cameras are not being recorded, and PTZ cameras change their settings) and a *dwell* (during which cameras capture, with constant settings, and are being recorded). Candidate plans differ in the number and duration of planning steps up to the planning horizon.

To ensure real-time performance, we decompose the optimization problem into subproblems and solve each subproblem independently. Our method consists of two components: centralized *global assignment* and distributed *local planning*.

The global assignment component accomplishes two tasks: grouping targets into *agencies* and assigning cameras to each agency. We create agencies by clustering together targets that are close to each other and predicted to be heading in the same direction. We use predicted target trajectories to cluster the targets into a minimum number of non-overlapping agencies of a given maximum diameter. We use a *minimal change* clustering heuristic that tries to preserve agency membership over time. We then use a greedy heuristic to assign cameras to each agency, based on their potential contribution to it. The heuristic iteratively tries assigning all available cameras to nearby agencies, searching for the camera-agency assignment that best improves the performance metric value for the agency. Improvement is measured using the ratio between the metric values before and after making the assignment. The resulting plans are compared with plans obtained by prolonging the current plans up to the planning horizon whenever possible, and the greedy assignments are only applied if they perform better. We use the same process both to control PTZ cameras in real-time and to select which cameras to record when there are fewer recording devices than cameras. In the case of selection, we simply stop after the maximum allowable number of cameras have been assigned. The plans corresponding to each camera-agency assignment are generated assuming the worst-case scenario: the camera is repeatedly set to transition, then capture for as long as possible, with PTZ cameras zoomed out to a field of view as wide as possible. Predicted static and dynamic occlusions are taken into account, and transitions are planned during occlusions whenever possible, in order to minimize the time intervals when cameras are not capturing.

Local planning at the level of each agency is concerned with the locally-optimal capture of the targets in the agency. All cameras assigned to each agency capture all member targets, and no further camera-target assignment decisions are made at this level. The planning decisions made at this level are on when and for how long each camera should dwell (capture), and when each PTZ camera should transition to a new configuration. All possible combinations of candidate plans for all cameras are explored exhaustively using backtracking. To achieve on-line, real-time control, the set of candidate plans is heuristically generated

and sorted so that the most promising plans are evaluated first. We use prior experimental observations to derive criteria for judging a plan's potential. While not a guarantee that the best plan would be chosen on time, we have found this heuristic to closely approximate an exhaustive search.

3.3 Experimental Results

We have applied ASCENT to automated on-line control of cameras in simulated and laboratory settings, capturing training exercises that involved patrolling, cordoning and searching a civilian, and crossing a danger zone. Experiments showed the emergence of desired camera behaviors, including: fast coverage of new targets, continuous target coverage via staggered settings adjustments, continuous coverage of divergent target groups, automatic hand-offs, and continuous preemptive coverage of fast-moving targets. The performance metric and control method were tuned to produce images best suited for a volumetric reconstruction method such as [7]. The interested reader is referred to Chapter 7 of [10] and [8] for more details.

The simulated setting involved capturing 6 targets (4 Marines and 2 civilians) moving around 2 occluders, using 6 cameras. Figure 1 (Left) shows an overview of the setup as modeled in the simulator. Camera locations are shown in blue, occluders are shown in red. The laboratory setting involved capturing 7 targets (4 Marines and 3 civilians) moving around the entrance to an alley between 2 buildings, using 8 cameras hanging from the ceiling. Figure 1 (Right) shows an image captured by an overview camera during the exercise. Building walls were simulated using cloth attached to waist-high posts.

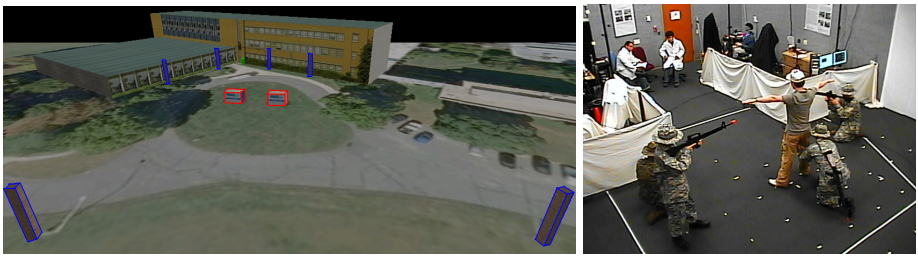

Fig. 1. (Left) Simulated setting. (Right) Laboratory setting.

4 Training Support Framework

We envision ASCENT as part of a training support framework that defines how automatic control of cameras can augment the capabilities at training ranges.

First, by automating camera selection and control decisions, ASCENT augments the operators' capabilities. A well-configured automated system can make decisions that an operator may find counter-intuitive, but are justified when the

captured images are destined for automated analysis, as opposed to manual review. We envision the following scenarios for how ASCENT can be applied to augment human decisions in camera selection and control:

1. *Automated:* ASCENT controls all active cameras and selects a number of cameras for recording.
2. *Directed:* ASCENT allows operators to intervene on-the-fly, and designate important events, areas and persons for capture with higher priority. Camera selection and control are still done automatically, but operator interventions are incorporated as constraints in the optimization method.
3. *Assisted:* ASCENT allows operators to dynamically choose a set of cameras that they want to record, control directly or assign to particular targets. It then assists the operators by automatically selecting which of the remaining cameras to record, and controlling the remaining active cameras. It also suggests the best camera-target assignments and camera settings for the cameras chosen by the operators, but lets the operators decide whether to apply them or not.

Second, ASCENT augments the training capabilities at training ranges by helping provide images best suited for automated computer vision analysis, which has the potential to shorten AARs video packages down to segments relevant for improving the trainees' performance. To that end, both components of ASCENT are highly customizable. The performance metric can be adapted to include performance factors relevant to the application, such as varying weights for different members of a team over time; or factors relevant to the computer vision algorithm used, such as preferred incidence angles for 3D reconstruction or 2D posture recognition. The selection and control method can incorporate domain knowledge such as the training range topology and the locations of important training events in relation to camera placement, as well as their timing during a training exercise. The interested reader can find a discussion of many of the customizations possible in ASCENT in [8] and Chapters 5 and 6 of [10].

5 Conclusions and Future Work

We introduced a framework for augmenting capabilities at training ranges. Its main component is ASCENT, an optimization-based on-line camera selection and control approach consisting of a performance metric and a selection and control method. For the optimization objective function, we employ a versatile performance metric that can incorporate both camera performance factors and application requirements. To reduce the size of the search space and arrive at an implementation that runs in real-time, our camera control method breaks down the optimization problem into subproblems. We first use a proximity-based minimal change heuristic to decompose the problem into subproblems and a greedy heuristic to select cameras and assign them to subproblems. We then solve each subproblem independently, generating and evaluating candidate plans as time allows. We applied ASCENT to simulated and laboratory settings, demonstrating

useful camera behaviors. We briefly discussed how ASCENT can help augment the capabilities at training ranges: it can automate selection and control decisions, and can be easily adapted to include requirements for automated analysis using computer vision approaches.

We are looking forward to applying ASCENT in training ranges that have the camera infrastructure already in place, and gather feedback from camera operators, instructors and trainees. We plan to address the challenges of scaling an approach that has only been tested in simulated and laboratory settings with a small number of cameras to training ranges with hundreds of cameras. We are also looking forward to incorporating the requirements of emerging approaches that go beyond the results of today's computer vision algorithms and extract relevant information such as the video segments best suited for AARs. While in its current version ASCENT can capture images best suited for computer vision, human reviewers may have different requirements for AAR. We plan to leverage the experience of human operators in selecting footage appropriate for AARs in further customizing ASCENT to incorporate these requirements. Similarly, the experience of instructors currently following monitoring exercises on the ground will be invaluable.

Acknowledgments. We acknowledge our sponsors and collaborators in the "Behavior Analysis and Synthesis for Intelligent Training (BASE-IT)" project: ONR grant N00014-08-C-0349, Roy Stripling, Ph.D., Program Manager, led by Amela Sadagic (PI) at the Naval Post-graduate School, Greg Welch (PI) at UNC, and Rakesh Kumar (PI) and Hui Cheng (Co-PI) at Sarnoff.

References

1. Allen, B.D.: Hardware Design Optimization for Human Motion Tracking Systems. Ph.D. thesis, University of North Carolina at Chapel Hill (December 2007)
2. Broaddus, C., Germano, T., Vandervalk, N., Divakaran, A., Wu, S., Sawhney, H.: Act-vision: active collaborative tracking for multiple ptz cameras. In: Proceedings of SPIE: Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications, Orlando, FL, USA, vol. 7345 (April 2009)
3. Denzler, J., Zobel, M., Niemann, H.: On optimal camera parameter selection in kalman filter based object tracking. In: 24th DAGM Symposium on Pattern Recognition, pp. 17–25 (2002)
4. Denzler, J., Zobel, M., Niemann, H.: Information theoretic focal length selection for real-time active 3-d object tracking. In: International Conference on Computer Vision, vol. 1, pp. 400–407 (October 2003)
5. Deutsch, B., Niemann, H., Denzler, J.: Multi-step active object tracking with entropy based optimal actions using the sequential kalman filter. In: IEEE International Conference on Image Processing, vol. 3, pp. 105–108 (2005)
6. Deutsch, B., Zobel, M., Denzler, J., Niemann, H.: Multi-step entropy based sensor control for visual object tracking. In: Rasmussen, C.E., Bülthoff, H.H., Schölkopf, B., Giese, M.A. (eds.) DAGM 2004. LNCS, vol. 3175, pp. 359–366. Springer, Heidelberg (2004)

7. Guan, L.: Multi-view Dynamic Scene Modeling. Ph.D. thesis, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA (April 2010)
8. Ilie, A., Welch, G.: On-line control of active camera networks for computer vision tasks. ACM Transactions on Sensor Networks (to appear, 2014)
9. Ilie, A., Welch, G., Macenko, M.: A stochastic quality metric for optimal control of active camera network configurations for 3D computer vision tasks. In: Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications, Marseille, France (October 2008)
10. Ilie, D.A.: On-Line Control of Active Camera Networks. Ph.D. thesis, University of North Carolina at Chapel Hill (2010)
11. Krahnstoever, N., Yu, T., Lim, S.N., Patwardhan, K., Tu, P.: Collaborative real-time control of active cameras in large scale surveillance systems. In: Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications, Marseille, France (October 2008)
12. Kutulakos, K.N., Seitz, S.M.: A theory of shape by space carving. International Journal of Computer Vision 38(3), 199–218 (2000)
13. Lim, S.-N., Davis, L., Mittal, A.: Task scheduling in large camera networks. In: Yagi, Y., Kang, S.B., Kweon, I.S., Zha, H. (eds.) ACCV 2007, Part I. LNCS, vol. 4843, pp. 397–407. Springer, Heidelberg (2007)
14. Matsuyama, T., Wu, X., Takai, T., Nobuhara, S.: Real-time 3d shape reconstruction, dynamic 3d mesh deformation, and high fidelity visualization for 3d video. In: Computer Vision and Image Understanding, vol. 96, pp. 393–434. lsevier Science Inc., New York (2004)
15. Matsuyama, T., Ukita, N.: Real-time multitarget tracking by a cooperative distributed vision system. Proceedings of the IEEE 90, 1137–1150 (2002)
16. Matusik, W., Buehler, C., Raskar, R., Gortler, S.J., McMillan, L.: Image-based visual hulls. In: ACM Siggraph, pp. 369–374 (2000)
17. Naish, M.D., Croft, E.A., Benhabib, B.: Coordinated dispatching of proximity sensors for the surveillance of manoeuvring targets. Robotics and Computer-Integrated Manufacturing 19(3), 283–299 (2003)
18. Qureshi, F., Terzopoulos, D.: Surveillance in virtual reality: System design and multi-camera control. In: Proceedings of Computer Vision and Pattern Recognition, pp. 1–8 (June 2007)
19. Qureshi, F.Z., Terzopoulos, D.: Towards intelligent camera networks: a virtual vision approach. In: 2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, Beijing, China, pp. 177–184 (October 2005)
20. Sadagic, A., Welch, G., Basu, C., Darken, C., Kumar, R., Fuchs, H., Cheng, H., Frahm, J.M., Kolsch, M., Rowe, N., Towles, H., Wachs, J., Lastra, A.: New generation of instrumented ranges: Enabling automated performance analysis. In: 2009 Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC-2009), Orlando, FL (2009)
21. Sommerlade, E., Reid, I.: Probabilistic surveillance with multiple active cameras. In: IEEE International Conference on Robotics and Automation (May 2010)
22. Wachs, J., Goshorn, D., Kolsch, M.: Recognizing human postures and poses in monocular still images. In: Intl. Conf. on Image Processing, Computer Vision, and Pattern Recognition, Las Vegas, NV, pp. 665–671 (2009)