Available online at www.sciencedirect.com
SCIENCE DIRECT®
Computer Vision and Image Understanding
Computer Vision and Image Understanding 100 (2005) 198–224
www.elsevier.com/locate/cviu

Selective perception policies for guiding sensing and computation in multimodal systems: A comparative analysis

Nuria Oliver *, EricHorvitz

Adaptive Systems and Interaction, Microsoft Research, Redmond, WA 98052, USA

Received 26 January 2004; accepted 23 December 2004
Available online 23 June 2005

Abstract

Intensive computations required for sensing and processing perceptual information can impose significant burdens on personal computer systems. We explore the use of value of information to guide sensing and analysis in automated behavior recognition systems, and highlight the role of such computations as a formal foundation for selective attention. We examine several different policies for selective perception in SEER, a multimodal system for recognizing office activity that relies on a layered Hidden Markov Model (LHMM) representation. We review our efforts to employ expected value of information (EVI) computations to limit sensing and analysis in a context-sensitive manner. We discuss an implementation of a one-step myopic EVI analysis and compare the results of using the myopic EVI with a heuristic sensing policy that makes observations at different frequencies. Both policies are then compared to a random perception policy, where sensors are selected at random. Finally, we discuss the sensitivity of ideal perceptual actions to preferences encoded in utility models about information value and the cost of sensing.

Keywords: Selective perception; Expected value of information; Automatic feature selection; Hidden Markov models; Office awareness; Multimodal interaction; Human behavior recognition

* Corresponding author. Fax: +1 425 936 7329.
E-mail addresses: nuria@microsoft.com (N. Oliver), horvitz@microsoft.com (E. Horvitz).
URL: http://research.microsoft.com/~nuria (N. Oliver).

1077-3142/$ - see front matter
doi:10.1016/j.cviu.2004.12.004

1. Introduction

Investigators have been long interested in the promise of performing automatic recognition of human behavior and intentions from observations. Successful recognition of human behavior enables compelling applications, including automated visual surveillance and multimodal human–computer interaction (HCI)—considering multiple streams of information about a user's behavior and the overall *context* of a situation to provide appropriate control and services. There has been progress on multiple fronts in recognizing human behavior and intentions. However, challenges remain for developing machinery that can provide rich, human-centric notions of context in a tractable manner. We address in this paper the computational burden associated with perceptual analysis.

Computation for visual and acoustical analyses has typically required a large portion—if not nearly all—of the total computational resources of personal computers that make use of such perceptual inferences. It is not surprising to find that there is little interest in invoking such perceptual services when they require a substantial portion of the available CPU time, significantly slowing down more primary applications that are supported or extended by the perceptual apparatus. Thus, we have pursued coherent strategies for automatically limiting in an automated manner the computational load of perceptual systems, in a similar way as biological systems do.

We shall focus in this paper on the control of perception in SEER, a probabilistic reasoning system that provides real-time interpretations of human activity in and around an office [1]. We have explored two strategies for sensor selection and sensor data processing in SEER.

The first approach is based on the use of decision-theoretic principles to guide perception, where we compute the expected value of information (EVI) of different subsets of observations in real-time on a frame by frame basis. This is a greedy, one-step lookahead approach to computing the next best set of observations to evaluate at each time step. We refer to this strategy as EVI-based perception.

The second approach to limiting the computational burden of perception centers on defining heuristic policies in terms of observational frequencies and duty cycles with which each feature, extracted from the sensors, is computed. We refer to this approach rate-based perception.

We compare the performance of the EVI-based and the rate-based perception methods with the legacy SEER system that analyzes all features all the time (i.e., without selective perception), and with a random feature selection perception approach, where the features are randomly selected at each time step.

This paper is organized as follows: we first provide background on context-sensing systems and principles for guiding perception in Section 2. In Section 3, we describe the challenge of understanding human activity in an office setting and review the different perceptual inputs that are used. We also provide background on the legacy SEER system, focusing on our work to extend a single-layer implementation of HMMs into a more effective cascade of HMMs, a representation that we refer to as Layered Hidden Markov Models (LHMMs). Section 4