



## IMPROVED $Q_{MDP}$ POLICY FOR PARTIALLY OBSERVABLE MARKOV DECISION PROCESSES IN LARGE DOMAINS: EMBEDDING EXPLORATION DYNAMICS

**GIORGOS APOSTOLIKAS AND SPYROS TZAFESTAS**

*Intelligent Robotics and Automation Laboratory*

*Department of Signals, Control and Robotics*

*School of Electrical and Computer Engineering*

*National Technical University of Athens*

*Zografou 15773, Athens, Greece*

*e-mail: tzafesta@sofilab.ntua.gr, zupelman@yahoo.com*

**ABSTRACT**—Artificial Intelligence techniques were primarily focused on domains in which at each time the state of the world is known to the system. Such domains can be modeled as a Markov Decision Process (MDP). Action and planning policies for MDPs have been studied extensively and several efficient methods exist. However, in real world problems pieces of information useful for the process of action selection are often missing. The theory of Partially Observable Markov Decision Processes (POMDP's) covers the problem domain in which the full state of the environment is not directly perceivable by the agent. Current algorithms for the exact solution of POMDP's are only applicable to domains with a small number of states. To cope with more extended state spaces, a number of methods that achieve sub-optimal solutions exist and among these the  $Q_{MDP}$  approach seems to be the best. We introduce a novel technique, called Explorative  $Q_{MDP}$  ( $EQ_{MDP}$ ) which constitutes an important enhancement of the  $Q_{MDP}$  method. To the best knowledge of the authors,  $EQ_{MDP}$  is currently the most efficient method applicable to large POMDP domains.

**Key Words:** POMDP,  $Q_{MDP}$ , Action Selection