



















## REFERENCES

- [1] J.-Y. Audibert, R. Munos, and C. Szepesvári. Exploration–exploitation tradeoff using variance estimates in multi-armed bandits. *Theoretical Computer Science*, 410(19):1876–1902, 2009.
- [2] P. Auer. Using confidence bounds for exploitation-exploration trade-offs. *The Journal of Machine Learning Research*, 3:397–422, 2003.
- [3] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2-3):235–256, 2002.
- [4] O. Bayazit, J.-M. Lien, and N. Amato. Better group behaviors in complex environments using global roadmaps. In *8th International Conference on Artificial life*, pages 362–370, 2003.
- [5] R. A. Brooks. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, RA-2(1):14–23, Mar. 1986.
- [6] P. Fiorini and Z. Shiller. Motion planning in dynamic environments using Velocity Obstacles. *International Journal of Robotics Research*, 17:760–772, 1998.
- [7] A. Garivier and E. Moulines. On upper-confidence bound policies for switching bandit problems. In *Algorithmic Learning Theory*, pages 174–188. Springer, 2011.
- [8] S. Guy, S. Kim, M. Lin, and D. Manocha. Simulating heterogeneous crowd behaviors using personality trait theory. In *ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 43–52, 2011.
- [9] D. Helbing and P. Molnar. Social force model for pedestrian dynamics. *Physical review E*, 51(5):4282, 1995.
- [10] D. Hsu, R. Kindel, J. Latombe, and S. Rock. Randomized kinodynamic motion planning with moving obstacles. *International Journal of Robotics Research*, 21(3):233–255, 2002.
- [11] I. Karamouzas and M. Overmars. Simulating and evaluating the local behavior of small pedestrian groups. *IEEE Transactions on Visualization and Computer Graphics*, 18(3):394–406, 2012.
- [12] O. Khatib. Real-time obstacle avoidance for manipulators and mobile robots. *International Journal of Robotics Research*, 5(1):90–98, 1986.
- [13] S. Kim, S. J. Guy, W. Liu, R. W. Lau, M. C. Lin, and D. Manocha. Predicting pedestrian trajectories using velocity-space reasoning. In *Algorithmic Foundations of Robotics X*, pages 609–623. Springer, 2013.
- [14] S. M. LaValle. *Planning algorithms*. Cambridge University Press, 2006.
- [15] W. G. Macready and D. H. Wolpert. Bandit problems and the exploration/exploitation tradeoff. *IEEE Transactions on Evolutionary Computation*, 2(1):2–22, 1998.
- [16] F. Martínez-Gil, M. Lozano, and F. Fernández. Multi-agent reinforcement learning for simulating pedestrian navigation. In *Adaptive and Learning Agents*, pages 54–69. Springer, 2012.
- [17] F. Martínez-Gil, M. Lozano, and F. Fernández. MARL-Ped: A multi-agent reinforcement learning based framework to simulate pedestrian groups. *Simulation Modelling Practice and Theory*, 47:259–275, 2014.
- [18] M. Nowak and K. Sigmund. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner’s dilemma game. *Nature*, 364(6432):56–58, 1993.
- [19] M. A. Nowak. Five rules for the evolution of cooperation. *Science*, 314(5805):1560–1563, 2006.
- [20] J. Ondřej, J. Pettré, A.-H. Olivier, and S. Donikian. A synthetic-vision based steering approach for crowd simulation. In *ACM Transactions on Graphics (TOG)*, volume 29, page 123. ACM, 2010.
- [21] N. Pelechano, J. Allbeck, and N. Badler. Controlling individual agents in high-density crowd simulation. In *ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 99–108, 2007.
- [22] H. Robbins. Some aspects of the sequential design of experiments. In *Herbert Robbins Selected Papers*, pages 169–177. Springer, 1985.
- [23] W. Shao and D. Terzopoulos. Autonomous pedestrians. *Graphical Models*, 69(5-6):246–274, 2007.
- [24] R. S. Sutton. Learning to predict by the methods of temporal differences. *Machine learning*, 3(1):9–44, 1988.
- [25] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 1998.
- [26] M. Tokic. Adaptive  $\epsilon$ -greedy exploration in reinforcement learning based on value differences. In *KI 2010: Advances in Artificial Intelligence*, pages 203–210. Springer, 2010.
- [27] L. Torrey. Crowd simulation via multi-agent reinforcement learning. In *Artificial Intelligence and Interactive Digital Entertainment*, pages 89–94, 2010.
- [28] J. van den Berg, S. J. Guy, M. Lin, and D. Manocha. Reciprocal n-body collision avoidance. In *Robotics Research: The 14th International Symposium ISRR*, volume 70 of *Springer Tracts in Advanced Robotics*, pages 3–19. Springer-Verlag, 2011.
- [29] C. Zhang and V. Lesser. Coordinated multi-agent learning for decentralized pomdps. In *7th Annual Workshop on Multiagent Sequential Decision-Making Under Uncertainty (MSDM) at AAMAS*, pages 72–78, 2012.
- [30] C. Zhang and V. Lesser. Coordinating multi-agent reinforcement learning with limited communication. In *12th International Conference on Autonomous Agents and Multi-Agent Systems*, pages 1101–1108, 2013.