This paper focuses on multi-person action forecasting in videos. More precisely, given a history of H previous frames, the goal is to detect actors and to predict t