**Sparseness Meets Deepness: 3D Human Pose Estimation from Monocular Video**

**The 7 W's**

Q1. This paper addresses the challenge of 3D full-body hu- man pose estimation from a monocular image sequence. Here, two cases are considered: (i) the image locations of the human joints are provided and (ii) the image locations of joints are unknown.

Q2. It is relevant because it achieves greater 3D accuracy over state-of-the-art models and also outperforms existing 2D pose estimation baselines.

Q3. Early research on 3D monocular pose estimation in videos largely centered on incremental frame-to-frame pose tracking, e.g., [8, 42, 38]. These approaches rely on a given pose and dynamic model to constrain the pose search space. Another strand of research has focused on methods that predict 3D poses by searching a database of exemplars [36, 27, 19] or via a discriminatively learned mapping from the image directly or image features to human joint locations [1, 34, 51, 16, 44].

Q4. Notable drawbacks of this approach include: the requirement that the initialization be provided and their inability to recover from tracking failures.

Q5. This paper presents a 3D pose recovery framework that consists of a novel synthesis between discriminative image- based and 3D reconstruction approaches. In particular, the approach reasons jointly about image-based 2D part location estimates and model-based 3D pose reconstruction, so that they can benefit from each other.

Q6. Empirical evaluation on the Human3.6M dataset shows that the proposed approaches achieve greater 3D pose estimation accuracy over state-of-the-art base- lines. Further, the proposed approach outperforms a publicly available 2D pose estimation baseline on the challenging PennAction dataset.