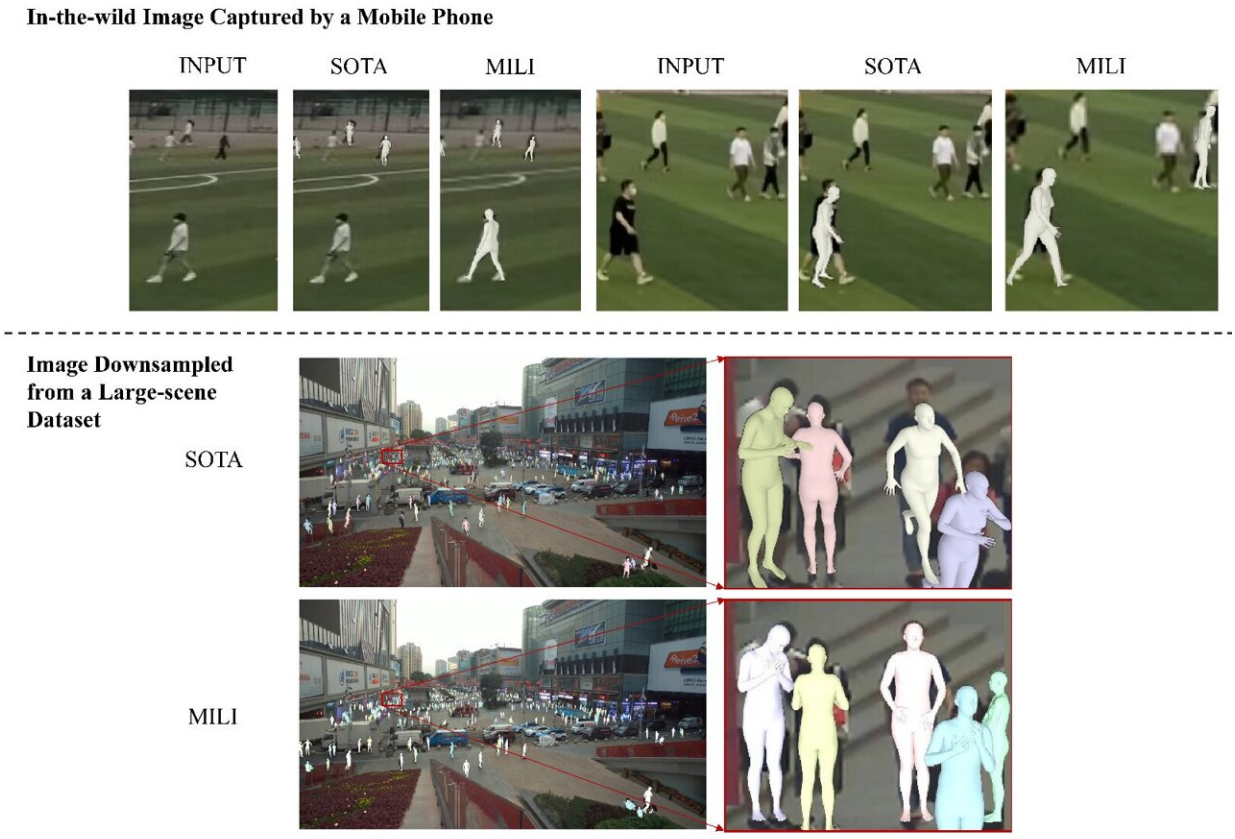
Estimation of multi-person 3D poses and shapes from a low-resolution image

April 17 2023



Using a low-resolution image captured by a mobile phone or down sampled from a large-scene dataset, the new method MILI (multi-person inference from a low-resolution image) can achieve more accurate multi-person reconstruction compared with a state-of-the-art (SOTA) method. Credit: The Authors

Accurately estimating 3D poses and body shapes from a single image is critical for several applications, such as behavior analysis and security alerts. Unfortunately, many existing multi-person reconstruction methods require the people present to be clearly visible in the photo to supply enough information. This becomes a problem when cameras have limited resolutions and the field of view is increased to capture individuals in distant areas, resulting in low-resolution images that provide little information.

To address that limitation, a research team from Tianjin University and Cardiff University attempted to reconcile the conflict between image resolution and estimation accuracy. As reported in the KeAi journal *Fundamental Research*, the team proposed an end-to-end multi-task machine learning framework known as MILI (multi-person inference from a low-resolution image) that enables accurate multi-person 3D pose and shape representation from a low-resolution image.

Further, to tackle the occlusion issue in multi-person scenes, the researchers devised an occlusion-aware mask prediction network for estimating the mask of each person's mesh during regression. Pair-wise images with high and low resolution were also used for training.

"In both small-scale and large-scale scenes, MILI outperformed the state-of-the-art methods both quantitatively and qualitatively," said Kun Li, lead author of the study. "Different from the existing work, MILI, as an end-to-end network, encourages the multi-person reconstruction even from low-resolution images and significantly improves the robustness to occlusions with the occlusion-aware mask prediction network by refining the detection stage with segmentation."

The code is available here.

"Reconstruction of 3D poses and shapes for the individuals in a

surveillance scene will allow for better recognition of actions/activities, including the interaction between people, modeling crowd behavior for simulations and security monitoring, and better tracking of individuals over time," concluded Li.

More information: Kun Li et al, MILI: Multi-person inference from a low-resolution image, *Fundamental Research* (2023). DOI: 10.1016/j.fmre.2023.02.006

Provided by KeAi Communications Co.