

Errata for the paper

Towards understanding action recognition

H, Jhuang, J. Gall, S. Zuffi, C. Schmid and M. J. Black
ICCV 2013

The version of the paper appearing on the conference USB stick, contains an error in our evaluation of Yang & Ramanan’s pose estimation algorithm [1]. After the re-evaluation, we found pose features derived from [1] already outperform flow-based features on a subset of our data where the full body is visible. We have updated results and conclusions accordingly. The on-line version of the paper is the corrected version; this is the official version of the paper.

For completeness, we summarize below changes from the version on the conference USB stick.

	In the old version	In the corrected version
Abstract	but current pose estimation algorithms are not yet reliable enough to provide this information.	While current pose estimation algorithms are far from perfect, features extracted from estimated pose on a subset of J-HMDB , in which the full body is visible, outperform low/mid-level features.
The last sentence in the second last paragraph in the Introduction	Unfortunately current pose estimation algorithms are not reliable enough for action recognition.	Remove to save space
The last paragraph in the Introduction	Our preliminary results show that pose features estimated from [33] still perform poorly on a realistic dataset like JHMDB	Our preliminary results show that pose features estimated from [33] perform much worse than the ground truth pose features, but they outperform low/mid level features for action recognition on clips where the full body is visible.
Table 3 (5)	14.0 40.0 34.1 31.0 31.8	19.9 45.6 49.8 54.1 52.9
The Fourth line in the last paragraph of section 6.2	Using the error measurement in [7], the pose estimation accuracy is 10%.	Using the error measurement in [7] with threshold 0.15, the pose estimation accuracy is 22.4%.
The seventh line in the last paragraph of section 6.2	results in a 11.9 pp drop ... results in a 31:0% accuracy for action recognition (Tab. 3 (5)).	results in a 3.8 pp gain ... results in a 8.1 pp gain over the baseline (Tab. 3 (5)).
The last sentence in the last paragraph of section 6.2	although high-level pose features outperform low/mid level features, one can not yet rely on current pose estimation algorithms to obtain this information.	while the estimated joint positions are not accurate compared to the ground truth, the derived pose features already outperform low/mid level features for action recognition.
The last sentence in the second last paragraph of section 7	the currently used pose estimation algorithm is not reliable enough to improve action recognition on	for sub-J-HMDB , where the full body is visible, a recent pose estimation algorithm computes poses that are more reliable than low/mid level features for action recognition of

[1] Y. Yang and D. Ramanan. Articulated human detection with flexible mixtures of parts. PAMI, to appear.