Introduction:
➢ Our goal is to capture human body motion under changing lighting conditions in a multiview setup.

Actor Model:
➢ We augment the highly simplified BlobTracker human model introduced by [Stoll et al.] with a textured mesh (automatically skinned to the skeleton) with labeled materials.

Pose Tracking:
➢ Goal: Run our augmented BlobTracker approach taking as input optimal illumination-invariant material segmentations (with influence $w_i$) as well as 2D joint detections ($v_{id}$) to robustly estimate the body motion.

➢ Key idea: design an iterative approach to alternatively estimate materials and body pose using temporal cues.

➢ Adaptive weighting ($w_i$, $w_j$): temporally measure the quality of material segmentations (e.g. abrupt changes) and scale down/up relevance for tracking accordingly.

Lighting-Invariant Segmentation:
➢ Goal: obtain temporally and spatially consistent material segmentations, which are invariant from background complexity and appearance changes due to light, to feed to [Stoll et al.].

➢ Graph-cut Energy: cost of assigning material label $i_t$ to pixel $i$, $\forall i \in I$ (each frame/view is solved independently):

$$E(f) = \sum_{t} \left[ E_i^p(f_t) + E_i^l(f_t) \right] + \sum_{ij} E_{ij}(f_t, f_j)$$

➢ Pose Costs: sample 50 random poses from a Gaussian distribution around the current pose prediction $\hat{P}$ based on previous $P^{t-1}$, $P^{t-2}$:

$$E_i^l(f_t) = 1 - H_{i_t}(i_t)$$

➢ Appearance Costs: Mahalanobis distance between pixels and labels:

$$E_i^p(f_t) = \langle \Phi(x_i) - \mu_f \rangle^T C_f^{-1} \langle \Phi(x_i) - \mu_f \rangle$$

Feature image $\Phi(x_i) = [\sin(x_{xy}), \cos(x_{xy}), s_{xy}]$

Background feature $\mu_{BC}(x_i) = [\Phi(x_i)^T, E_i^l(x_{t-1}), \ldots, E_i^l(x_{t-1})]$

Material geometric median $\mu_f$ and covariance $C_f$ on the pose predicted locations $X_f = [x_i|e_i|e_i|e_i|e_i] > 1$

$$\hat{\mu_f} = \arg \min_{\mu_f} \sum_{y} \exp \left( \frac{-1}{2} \sum_{x_y} \langle x - \mu_f \rangle^T C_f^{-1} \langle x - \mu_f \rangle \right)$$

Smoothness: neighboring pixels with similar color have similar materials:

$$E_{ij}(f_t, f_j) = \exp \left( \frac{1}{2} \sum_{x_y} \langle x - \mu_f \rangle^T C_f^{-1} \langle x - \mu_f \rangle \right) \min(1, |f_t - f_j|)$$

Results:
➢ Our quantitative and qualitative results evidence that our approach accurately tracks the human pose and outperforms the existing methods.