Optimizing Over a Set of Manifolds*

Juergen Gall[1], Angela Yao[1], and Luc Van Gool[1,2]

[1] Computer Vision Laboratory, ETH Zurich, Switzerland
{gall,yaoa,vangool}@vision.ee.ethz.ch
[2] KU Leuven, Belgium

Abstract. We give a comprehensive description of the algorithm proposed in "2D Action Recognition Serves 3D Human Pose Estimation" [1].

1 Preliminaries

Having a skeleton and a surface model of the human, the human pose is represented by a vector in a bounded, high-dimensional state space $\mathbb{E} \subset \mathbb{R}^{D+6}$. While $\Theta = \theta_1, \cdots, \theta_D \in \mathbb{E}_\Theta$ denotes the joint angles, the global orientation and position are encoded by the 6D vector (r, t). An element of the search space is given by $x = (r, t, \Theta)$. We formulate pose estimation as an optimization problem over $\mathbb{E}$ for a given positive energy function V, i.e. $\min_{x \in \mathbb{E}} V(x)$. The energy function measures the consistency between the image and the projected surface of the human for a given pose x.

2 Baseline

As a baseline, we implemented the particle-based annealing optimization scheme ISA over $\mathbb{E}$ (Algorithm 1), which has been used in the multi-layer framework [2]. The optimization scheme, based on the theory of Feynman-Kac models [3], iterates over a selection and mutation step, and is also the underlying principle of the annealed particle filter [4]. In our experiments, we use the polynomial annealing scheme:

$$\beta_k = (k+1)^b \tag{1}$$

with $b = 0.7$. The mutation step is implemented with the scaling factor $\alpha_\Sigma = 0.4$ and the positive constant $\rho = 0.0001$. The set of particles is denoted by $\mathcal{S}$. An estimate of the pose is given by the weighted mean of the particles after the last iteration, i.e. $\hat{x} = \sum_{s^i \in \mathcal{S}} (w^i \cdot x^i)$. For r^i, the mean is computed in the space of rotations. The uniform distribution and the normal distribution are denoted by $\mathcal{U}[0, 1]$ and $\mathcal{N}(\mu, \Sigma)$, respectively.

* Technical Report 271, Computer Vision Laboratory, ETH Zurich. This work has been supported by funding from the Swiss National Foundation NCCR project IM2 as well as the EC project IURO. Angela Yao was also supported by funding from NSERC Canada.

Algorithm 1 Interacting Simulated Annealing over $\mathbb{E}$

For $k = 1, \ldots, It$
- *Selection*
  - $\forall s^i \in \mathcal{S}_{k-1}$: $w^i = \exp\left(-\beta_k \cdot V\left(r^i, t^i, \Theta^i\right)\right)$
  - $\forall s^i \in \mathcal{S}_{k-1}$: $w^i = w^i / \sum_{s^j \in \mathcal{S}_{k-1}} w^j$
  - $\mathcal{S}_k = \emptyset$; $\forall s^i \in \mathcal{S}_{k-1}$ draw u from $\mathcal{U}[0,1]$:

    If $u \leq w^i / \max_{s^j \in \mathcal{S}_{k-1}} w^j$ then
    - $\mathcal{S}_k = \mathcal{S}_k \cup \{s^i\}$

    Otherwise
    - $\mathcal{S}_k = \mathcal{S}_k \cup \{s^j\}$, where s^j is selected with probability w^j
- *Mutation*
  - $\mu = \frac{1}{|\mathcal{S}_k|} \sum_{s^j \in \mathcal{S}_k} (r^j, t^j, \Theta^j)$

    $\Sigma = \frac{\alpha \Sigma}{|\mathcal{S}_k|-1} \left(\rho I + \sum_{s^j \in \mathcal{S}_k} \left((r^j, t^j, \Theta^j) - \mu\right) \left((r^j, t^j, \Theta^j) - \mu\right)^T\right)$
  - $\forall s^i \in \mathcal{S}_k$ sample (r^i, t^i, Θ^i) from $\mathcal{N}((r^i, t^i, \Theta^i), \Sigma)$

3 Proposed Algorithm

We modify the baseline algorithm to optimize over a set of manifolds instead of a single state space. To this end, we consider a set of action classes $\mathcal{A} = \{a_1, \cdots, a_{|\mathcal{A}|}\}$, where we learn for each class an action-specific low-dimensional manifold $\mathbb{M}_a \subset \mathbb{R}^{d_a}$ with $d_a \ll D$. We assume that the following mappings are available:

$$f_a : \mathbb{E}_\Theta \mapsto \mathbb{M}_a, \quad g_a : \mathbb{M}_a \mapsto \mathbb{E}_\Theta, \quad h_a : \mathbb{M}_a \mapsto \mathbb{M}_a, \tag{2}$$

where f_a denotes the mapping from the state space to the low-dimensional manifolds, g_a the projection back to the state space, and h_a the prediction within an action-specific manifold. Since the manifolds encode only the space of joint angles, a low-dimensional representation of the full pose is denoted by $y_a = (r, t, \Theta_a)$ with $\Theta_a = f_a(\Theta)$. A particle $s^i = (y^i_a, a^i)$ stores the corresponding manifold label a^i in addition to the vector $y^i_a = (r^i, t^i, \Theta^i_a)$. While *Select* p_1 is outlined in Algorithm 2, *Optimization A*, *Select* p_2, and *Optimization B* are described in Algorithm 3. The particles in the manifolds $\mathbb{M}_a$ after *Optimization A* are denoted by $\mathcal{S}^{\mathbb{M}}_{It_A}$ and the particles in the state space after *Optimization B* are denoted by $\mathcal{S}^{\mathbb{E}}_{It_B}$. The probability of an action class a for a given frame t is denoted by $p(A = a \,|\, T = t, \mathcal{I})$ and the estimated joint angles of the previous frame are denoted by $\hat{\Theta}_{t-1}$.

Algorithm 2 *Select* p_1

- $\mathcal{S}^{\mathbb{M}} = \emptyset$; $\forall s^i \in \mathcal{S}^{\mathbb{M}}_{It_A}$ draw u from $\mathcal{U}[0,1]$:

  If $u < p_1$ then

  - $\mathcal{S}^{\mathbb{M}} = \mathcal{S}^{\mathbb{M}} \cup \{s^i\}$

  Otherwise

  - $\mathcal{S}^{\mathbb{M}} = \mathcal{S}^{\mathbb{M}} \cup \left\{\left(r^j, t^j, f_{a^j}(\Theta^j), a^j\right)\right\}$, where $(r^j, t^j, \Theta^j) \in \mathcal{S}^{\mathbb{E}}_{It_B}$ and a^j is selected with probability $p(A \mid T = t, \mathcal{I})$

References

1. Gall, J., Yao, A., van Gool, L.: 2d action recognition serves 3d human pose estimation. In: ECCV. (2010)
2. Gall, J., Rosenhahn, B., Brox, T., Seidel, H.P.: Optimization and filtering for human motion capture – a multi-layer framework. IJCV **87** (2010) 75–92
3. Moral, P.D.: Feynman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications. Springer, New York (2004)
4. Deutscher, J., Reid, I.: Articulated body motion capture by stochastic search. IJCV **61** (2005) 185–205

Algorithm 3 Optimizing over $\mathbb{M}_a$

Optimization A:

For $k = 1, \ldots, It_A$
- *Selection*
  - $\forall s^i \in \mathcal{S}^{\mathbb{M}}_{k-1}$: $w^i = \exp\left(-\beta_k \cdot V\left(r^i, t^i, g_{a^i}(\Theta^i_a)\right)\right)$
  - $\forall s^i \in \mathcal{S}^{\mathbb{M}}_{k-1}$: $w^i = w^i / \sum_{s^j \in \mathcal{S}^{\mathbb{M}}_{k-1}} w^j$
  - $\mathcal{S}^{\mathbb{M}}_k = \emptyset$; $\forall s^i \in \mathcal{S}^{\mathbb{M}}_{k-1}$ draw u from $\mathcal{U}[0,1]$:
    If $u \leq w^i / \max_{s^j \in \mathcal{S}^{\mathbb{M}}_{k-1}} w^j$ then
    - $\mathcal{S}^{\mathbb{M}}_k = \mathcal{S}^{\mathbb{M}}_k \cup \{s^i\}$

    Otherwise
    - $\mathcal{S}^{\mathbb{M}}_k = \mathcal{S}^{\mathbb{M}}_k \cup \{s^j\}$, where s^j is selected with probability w^j
- *Mutation*
  - $\forall a \in \mathcal{A}$: $\mu_a = \frac{1}{|\mathcal{S}_a|} \sum_{s^j \in \mathcal{S}_a} \Theta^j_a$ with $\mathcal{S}_a = \{s^i \in \mathcal{S}^{\mathbb{M}}_k : a^i = a\}$
    $\forall a \in \mathcal{A}$: $\Sigma_a = \frac{\alpha_\Sigma}{|\mathcal{S}_a|-1}\left(\rho\, I + \sum_{s^j \in \mathcal{S}_a} (\Theta^j_a - \mu_a)\,(\Theta^j_a - \mu_a)^T\right)$
    $\mu_0 = \frac{1}{|\mathcal{S}^{\mathbb{M}}_k|} \sum_{s^j \in \mathcal{S}^{\mathbb{M}}_k} (r^j, t^j)$
    $\Sigma_0 = \frac{\alpha_\Sigma}{|\mathcal{S}^{\mathbb{M}}_k|-1}\left(\rho\, I + \sum_{s^j \in \mathcal{S}^{\mathbb{M}}_k} \left((r^j, t^j) - \mu_0\right)\,\left((r^j, t^j) - \mu_0\right)^T\right)$
  - $\forall s^i \in \mathcal{S}^{\mathbb{M}}_k$ sample Θ^i_a from $\mathcal{N}(\Theta^i_a, \Sigma_{a^i})$ and (r^i, t^i) from $\mathcal{N}((r^i, t^i), \Sigma_0)$

Select p_2:

- $\hat{a} = \operatorname{argmin}_{a \in \mathcal{A}} \left\|\hat{\Theta}_{t-1} - g_a(f_a(\hat{\Theta}_{t-1}))\right\|$, $\quad (\Sigma_{\hat{a}})_{ii} = \frac{|\hat{\Theta}_{t-1} - g_{\hat{a}}(f_{\hat{a}}(\hat{\Theta}_{t-1}))|_i}{3}$
- $\mathcal{S}^{\mathbb{E}}_{It_A} = \emptyset$; $\forall s^i \in \mathcal{S}^{\mathbb{M}}_{It_A}$ draw u from $\mathcal{U}[0,1]$:
  If $u < p_2$ then
  - $\mathcal{S}^{\mathbb{E}}_{It_A} = \mathcal{S}^{\mathbb{E}}_{It_A} \cup \left\{\left(r^i, t^i, g_{a^i}(\Theta^i_a)\right)\right\}$

  Otherwise
  - $\mathcal{S}^{\mathbb{E}}_{It_A} = \mathcal{S}^{\mathbb{E}}_{It_A} \cup \left\{(r^i, t^i, \hat{\Theta})\right\}$, where $\hat{\Theta}$ is sampled from $\mathcal{N}(\hat{\Theta}_{t-1}, \Sigma_{\hat{a}})$

Optimization B:

For $k = It_A + 1, \ldots, It_B$
- *Selection*
  - $\forall s^i \in \mathcal{S}^{\mathbb{E}}_{k-1}$: $w^i = \exp\left(-\beta_k \cdot V\left(r^i, t^i, \Theta^i\right)\right)$
  - $\forall s^i \in \mathcal{S}^{\mathbb{E}}_{k-1}$: $w^i = w^i / \sum_{s^j \in \mathcal{S}^{\mathbb{E}}_{k-1}} w^j$
  - $\mathcal{S}^{\mathbb{E}}_k = \emptyset$; $\forall s^i \in \mathcal{S}^{\mathbb{E}}_{k-1}$ draw u from $\mathcal{U}[0,1]$:
    If $u \leq w^i / \max_{s^j \in \mathcal{S}^{\mathbb{E}}_{k-1}} w^j$ then
    - $\mathcal{S}^{\mathbb{E}}_k = \mathcal{S}^{\mathbb{E}}_k \cup \{s^i\}$

    Otherwise
    - $\mathcal{S}^{\mathbb{E}}_k = \mathcal{S}^{\mathbb{E}}_k \cup \{s^j\}$, where s^j is selected with probability w^j
- *Mutation*
  - $\mu = \frac{1}{|\mathcal{S}^{\mathbb{E}}_k|} \sum_{s^j \in \mathcal{S}^{\mathbb{E}}_k} (r^j, t^j, \Theta^j)$
    $\Sigma = \frac{\alpha_\Sigma}{|\mathcal{S}^{\mathbb{E}}_k|-1}\left(\rho\, I + \sum_{s^j \in \mathcal{S}^{\mathbb{E}}_k} \left((r^j, t^j, \Theta^j) - \mu\right)\,\left((r^j, t^j, \Theta^j) - \mu\right)^T\right)$
  - $\forall s^i \in \mathcal{S}^{\mathbb{E}}_k$ sample (r^i, t^i, Θ^i) from $\mathcal{N}((r^i, t^i, \Theta^i), \Sigma)$