TEMPORALLY CONSISTENT 3D POSE ESTIMATION IN THE INTERVENTIONAL ROOM USING MRF OPTIMIZATION OVER RGB-D SEQUENCES

Kadkhodamohammadi A.[1], Gangi A.[2], de Mathelin M.[1], and Padoy N.[1]

[1] ICube, University of Strasbourg, CNRS, France

[2] Radiology Department, University Hospital of Strasbourg, France

Abstract

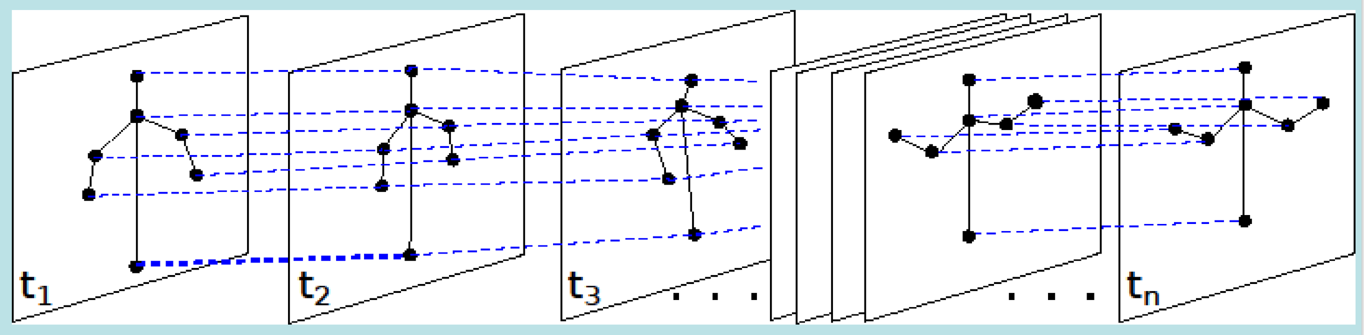
Tracking and estimating the pose of clinicians benefit many applications. However the special requirements of the operating room make the vision based tracking very challenging. In this paper, we propose an MRF energy formulation that leverages both kinematic and temporal constraints to estimate poses over RGB-D sequences. The quantitative evaluation of the approach on seven sequences from two different interventional rooms shows robust results in presence of multiple persons and occlusions.

Introduction

The use of interventional X-ray imaging has recently grown rapidly and therefore the risk of radiation exposure by clinician has increased dramatically. [1] shows that the amount of radiation absorption differs per body parts and a single dosimeter cannot provide a complete picture of radiation exposure. Since using several dosimeters is not practical, a vision-based system is needed for consistent tracking and estimation of body poses to accumulate radiation during the short bursts of emission from the X-ray device. Furthermore, such a system benefits many applications like workflow analysis and activity recognition. However, the operating room is a very challenging environment due to frequent illumination changes, similar colors of the clinicians' scrubs, high occlusion and limitations on camera positioning. Therefore we propose an MRF-based approach to consistently estimate clinicians' poses over entire sequences.

Method

We use an RGB-D sensor to capture the interventional room. The detector [2] is used as body part detector (BPD). We define an energy function to estimate poses in RGB-D sequences that enforces both body kinematic and temporal consistency.



MRF graph with kinematic and temporal edges

$$E(D) = \sum_{p \in P} V_p(d_p) + \lambda^k \sum_{(p,q)\in \epsilon^k} V^k_{p,q}(d_p, d_q) + \lambda^t \sum_{(p,q)\in \epsilon^t} V^t_{p,q}(d_p, d_q)$$

Data term incorporates image likelihood based on body part detector response defined as:

$$V_p = \begin{cases} M\left(C(d_p)\right) & if\ \#blobs(frame(p), label(p)) > 0 \\ \beta & otherwise \end{cases}$$

where $M(x) = \frac{x^2}{x^2+\alpha^2}$, and

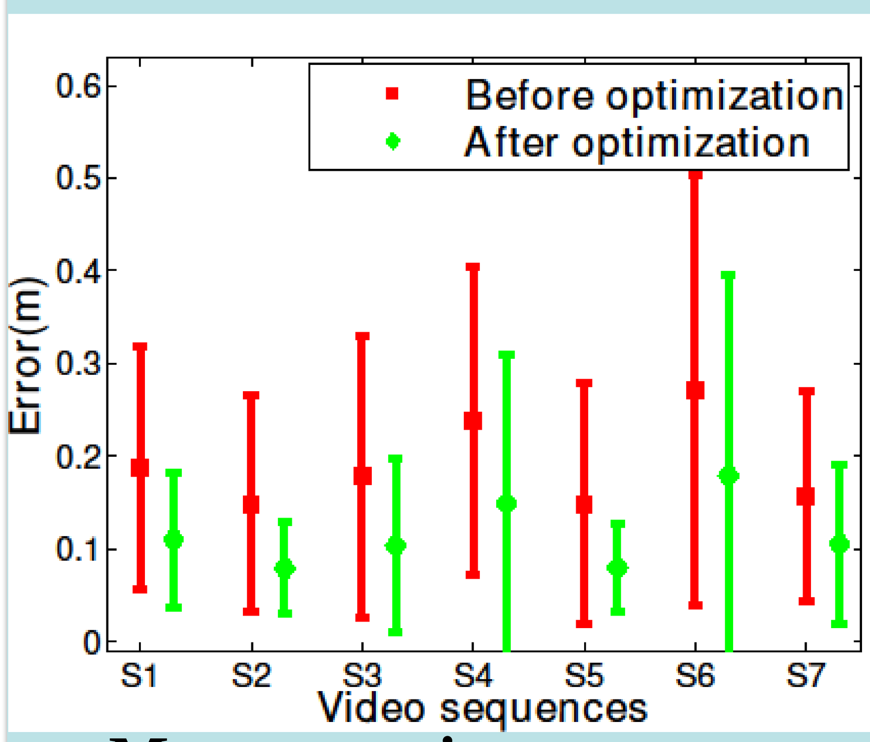
$$C(d_p) = \min_{b \in blobs(frame(p), label(p))} \parallel P(d_p) - centroid(b) \parallel * \left(\gamma - conf(b)\right).$$

Kinematic and temporal term are defined as:

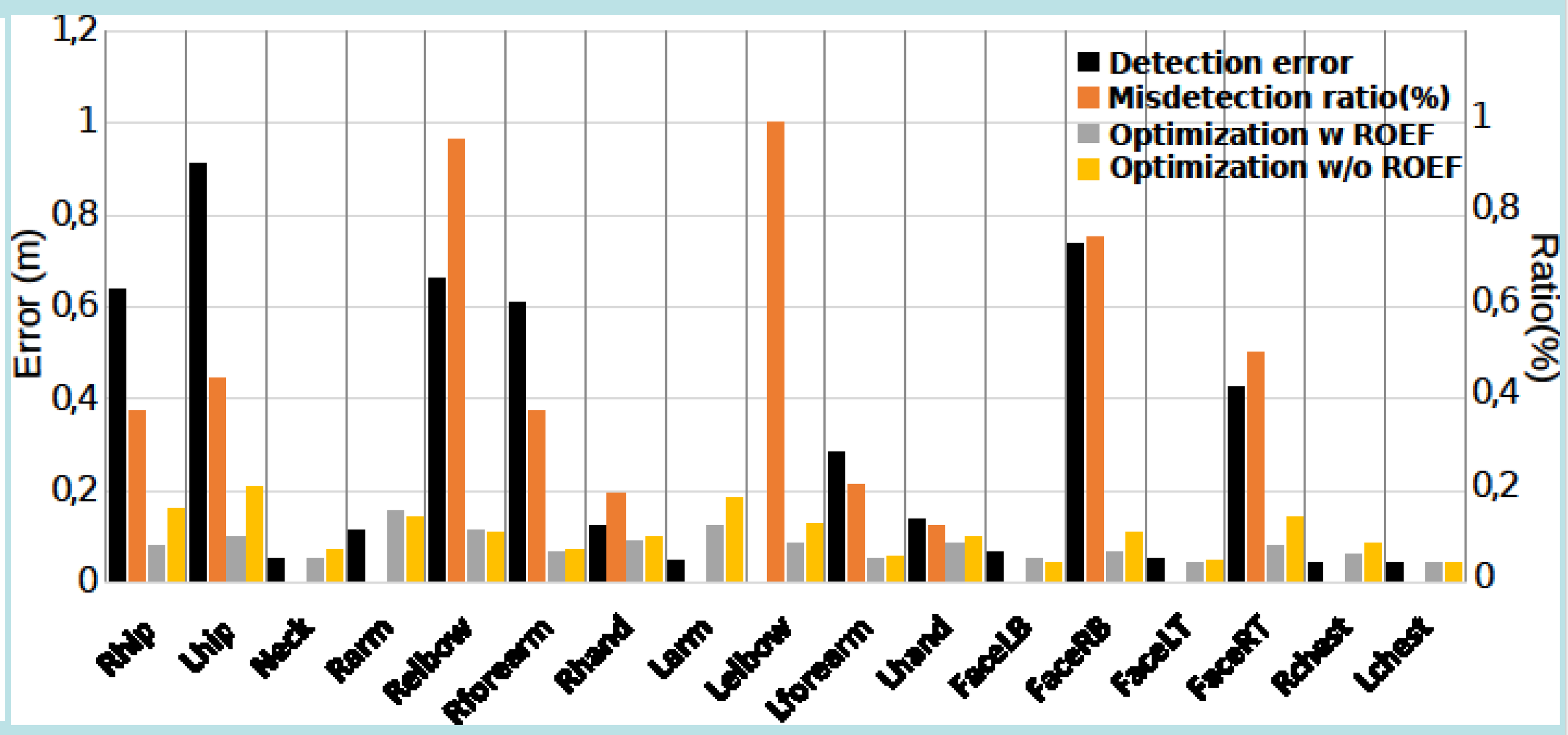
$$V^k_{p,q}(d_p, d_q) = |\parallel P(d_p) - P(d_q) \parallel - \mu_{pq}| \quad and \quad V^t_{p,q}(d_p, d_q) = \parallel P(d_p) - P(d_q) \parallel.$$

Results

The proposed approach is evaluated on seven manually annotated sequences recorded in two different interventional rooms. Two to three persons are present in the sequences.

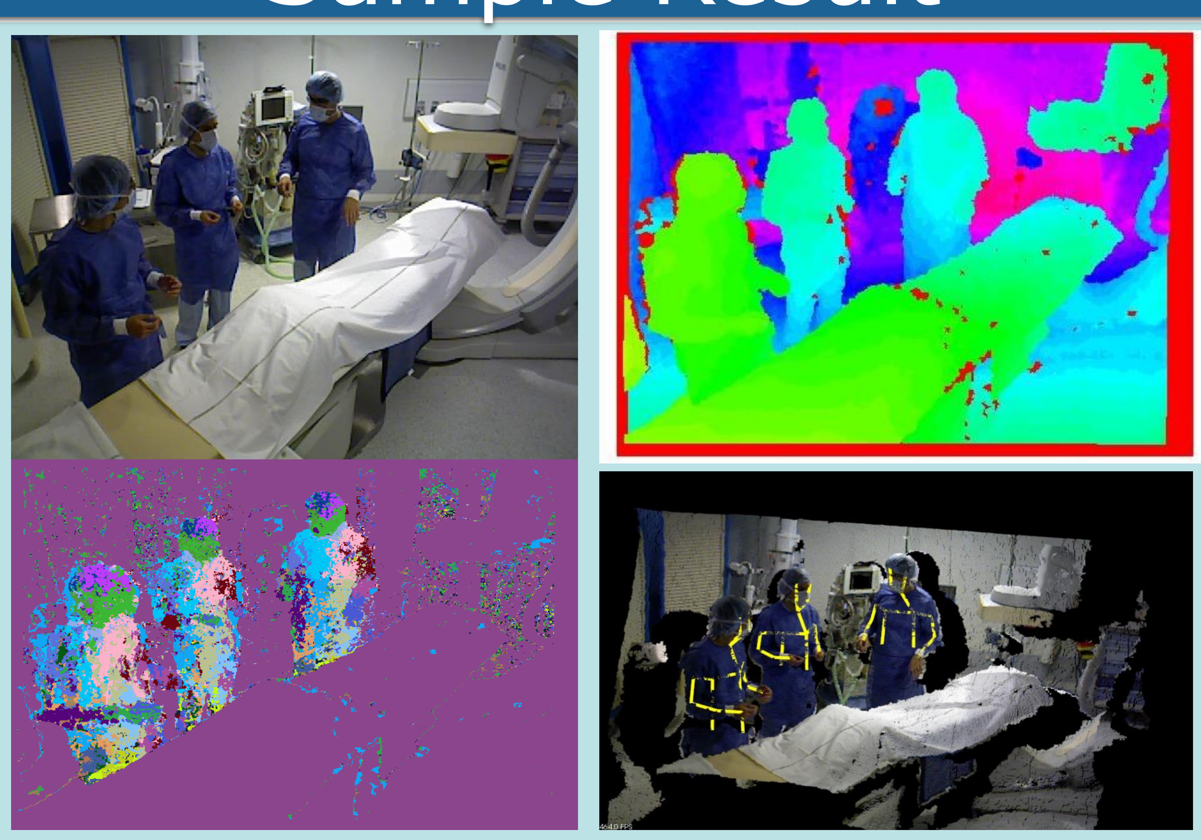


Mean error in meter per sequence at initialization and after optimization



Mean error in meter per body part for sequence S2 along with BPD misdetection

Sample Result



Input RGB and depth image, body part detector response and 3D skeleton overlaid on the reconstructed point cloud

Noisy Initialization

	ID	Initial	Optim.
Right hip	S1	1.02±0.36	0.11±0.05
All parts	S1	1.86±1.17	0.32±0.31
Right hip	S2	0.91±0.36	0.16±0.14
All parts	S2	1.81±1.21	0.31±0.33
Right hip	S3	0.94±0.38	0.13±0.13
All parts	S3	1.87±1.25	0.36±0.38

Mean error in meter with std before and after opt. for right hip and all parts

Conclusion

We propose an approach to track consistently the upper-body parts of clinical staff in an interventional room (IR) over RGB-D sequences. Our approach uses discrete optimization in an MRF framework that incorporate both kinematic and temporal constraints. The approach is quantitatively evaluated on seven manually-annotated RGB-D sequences recorded in two different IRs. The experiments show promising results in presence of multiple persons and occlusions. The experiments also show that the approach can recover from noisy initialization.

Ref.

[1] E., Carinou, et al., Recommendations to reduce extremity and eye lens doses in interventional radiology and cardiology. J. Rad. Meas., 2011.

[2] K., Buys, et al., An adaptable system for RGB-D based human body detection and pose estimation. J. Vis. Commun. Image R., 2013.