Inter-Camera Tracking Based On Fully Unsupervised Online Learning
Online-Learning-Based Human Tracking Across Non-Overlapping Cameras

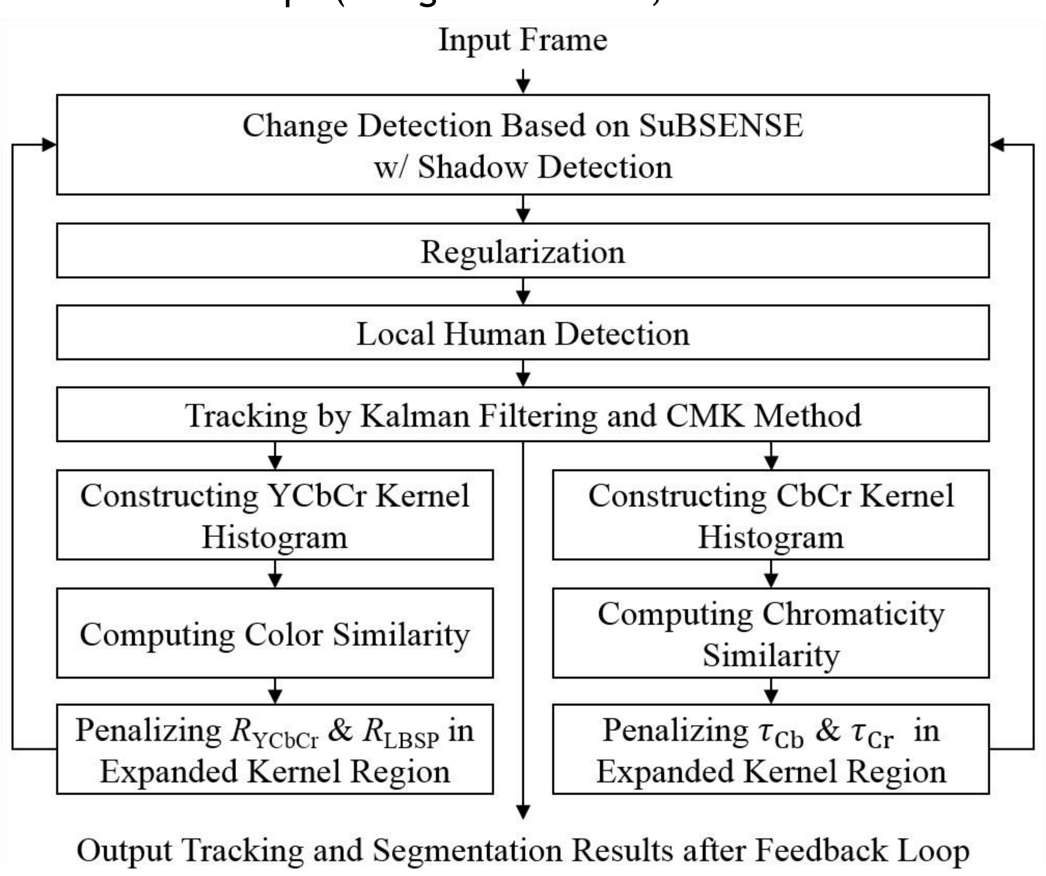
Young-Gun Lee[1], Zheng Tang[1], Jenq-Neng Hwang[1], Zhijun Fang[2]
[1]University of Washington, Seattle, WA 98195, USA, [2]Shanghai University of Engineering Science, Shanghai, China

Abstract

Due to the expanding scale of camera networks, multiple camera tracking of human has received higher attention in recent years. In this paper, we present a novel approach to track each human within a single camera and across multiple disjoint cameras. Our framework includes a multi-object tracking and segmentation system, a two-phase feature extractor, and an online-learning-based camera link model estimation. For tracking within a single camera, we apply tracking by segmentation and local object detection with multi-kernel feedback to adaptively improve robustness of the algorithm. In inter-camera tracking, we introduce an effective integration of appearance and context features. Automatically couples are detected, and the couple feature is also integrated with existing features. The proposed algorithm is scalable by a fully unsupervised online learning framework. In our experiments, the proposed method outperforms all the state-of-the-art in the benchmark NLPR MCT dataset.

Single-Camera Tracking and Object Segmentation

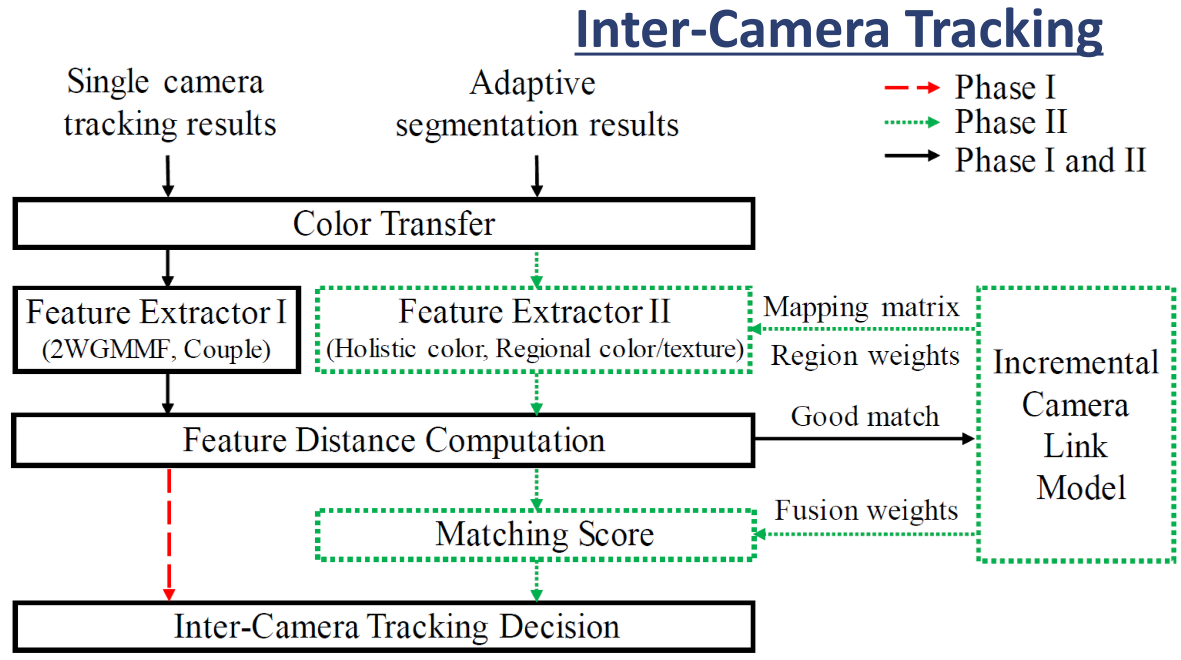
- Flow diagram of Multi-kernel Adaptive Segmentation and Tracking for SCT & segmentation.
- Comparison of segmentation performance. (a) Segmentation from the preliminary result of SuBSENSE with shadow detection. (b) Segmentation after the application of multi-kernel feedback loops (foreground in red, and detected shadow in blue).




(a)


(b)

Inter-Camera Tracking



Color transfer

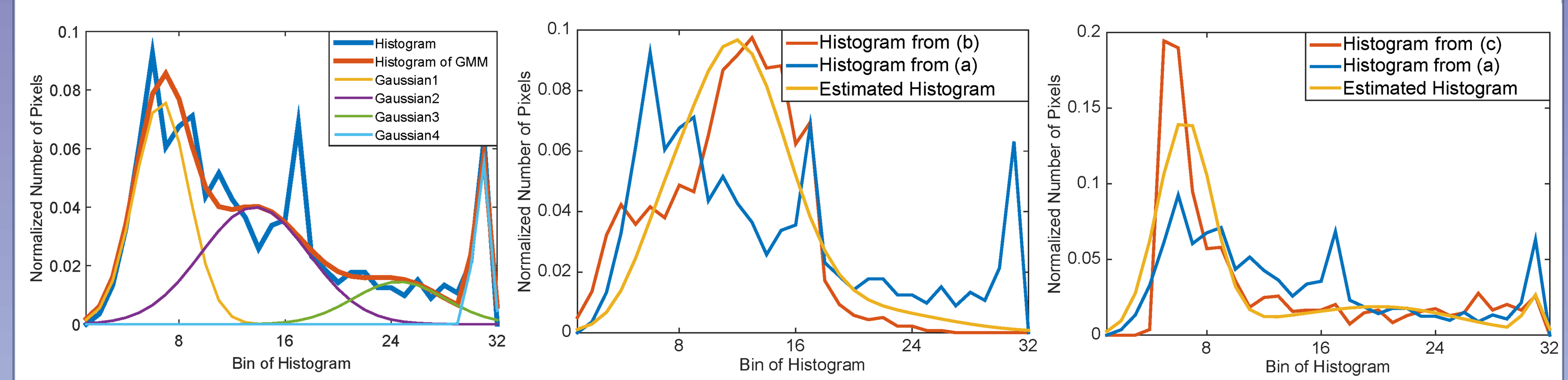
- RGB color space is transformed to the $la\beta$ color space and the data points composing the synthetic image are scaled by factors determined by the respective standard deviation.
- $I'_s = \frac{\sigma_t}{\sigma_s}(I_s - \mu_s) + \mu_t$.

Two-way Gaussian mixture model fitting feature

- Main idea of 2WGMMF feature is that main color modes of the same identity in color histogram should be consistent across different viewpoints.


(a) (b) (c)

$$d_{NL}(\mathbf{h}_i^A, G(\mathbf{h}_i^B)) = -\ln p(\mathbf{h}_i^A|\theta_1^B, \dots, \theta_K^B)$$
$$= -\ln\left(\sum_{k=1}^{K} \pi_k^B \mathcal{N}(\mathbf{h}_i^A|\boldsymbol{\mu}_k^B, \boldsymbol{\Sigma}_k^B)\right),$$



- Feature distance: $d_{\mathrm{2WGMMF}}(A,B) = d_{NL}(\mathbf{h}_{\mathrm{torso}}^A, G(\mathbf{h}_{\mathrm{torso}}^B)) + d_{NL}(\mathbf{h}_{\mathrm{legs}}^A, G(\mathbf{h}_{\mathrm{legs}}^B))$
$+ d_{NL}(\mathbf{h}_{\mathrm{torso}}^B, G(\mathbf{h}_{\mathrm{torso}}^A)) + d_{NL}(\mathbf{h}_{\mathrm{legs}}^B, G(\mathbf{h}_{\mathrm{legs}}^A))$.

Regional color and texture features

- The torso part is divided into six regions based on the pre-defined ratios.
- Since a specific region covers different areas of the torso due to different viewpoints, the histogram extracted from one region of the torso can be modeled as a linear combination of the histograms extracted from multiple regions of the torso in the other camera.

$$d_{\mathrm{regional\ feature}}(A,B) = \sum_{k=1}^{6} q_k \times d(\mathbf{h}_{map_k}^A, \mathbf{h}_{r_k}^B) + q_7 \times d(\mathbf{h}_{r_7}^A, \mathbf{h}_{r_7}^B),$$

where $\mathbf{h}_{map_k}^A = [\mathbf{h}_{r_1}^A \dots \mathbf{h}_{r_6}^A]\mathbf{w}_k$.

Couple feature

- A couple is defined as a pair of person traveling together through an FOV.
- After identifying the same couple across cameras, persons are re-identified.
- $d_{\mathrm{couple\ identifier}}(AC,BD) = \min\left(d_{\mathrm{2WGMMF}}(A,B), d_{\mathrm{2WGMMF}}(A,D)\right)$
$+ \min\left(d_{\mathrm{2WGMMF}}(C,B), d_{\mathrm{2WGMMF}}(C,D)\right)$.
- Person-to-person match in a couple:

$$d_{\mathrm{couple}}^{\mathrm{I}}(A,B) = -d_{\mathrm{2WGMMF}}\left(A, B_{\mathrm{couple}}\right) = -d_{\mathrm{2WGMMF}}(A,D)$$
$$d_{\mathrm{couple}}^{\mathrm{II}}(A,B) = -\sum_{j=1}^{N} \alpha_j d_{feature_j}^{Norm}(A,D)$$

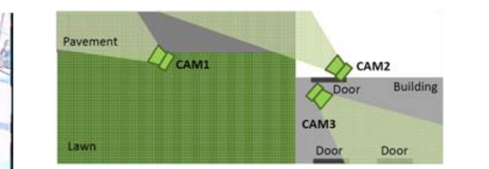
Final score

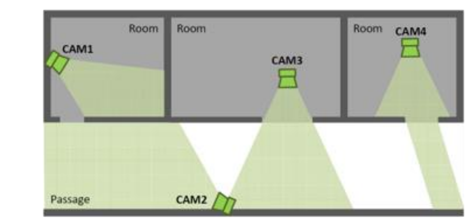
- Since the value range of each feature distance is different, min-max normalization and fusion methods are exploited to get the final score.
- $d_{\mathrm{Final}}^{\mathrm{I}}(A,B) = d_{\mathrm{2WGMMF}}(A,B) + d_{\mathrm{couple}}^{\mathrm{I}}(A,B)$, $\quad d_{\mathrm{Final}}^{\mathrm{II}}(A,B) = \sum_{j=1}^{N} \alpha_j d_{feature_j}^{Norm}(A,B)$,

where $d_j = \mu_j^{\mathrm{N}} - \mu_j^{\mathrm{P}} / \sqrt{\left(\sigma_j^{\mathrm{N}}\right)^2 + \left(\sigma_j^{\mathrm{P}}\right)^2}$ $\quad \alpha_j = d_j / \sum_{i=1}^{4} d_i$

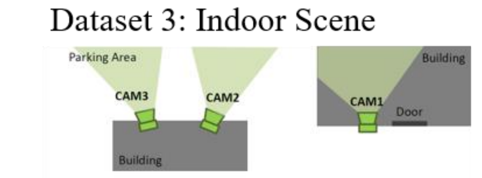
Dataset and evaluation criteria

Dataset 1-2: Outdoor and Indoor Scene


Dataset 1-2: Outdoor/Indoor Scene


Dataset 3: Indoor Scene

Dataset 3: Indoor Scene



Dataset 4: Outdoor Scene


Dataset 4: Outdoor Scene

- Evaluation criteria: $\mathrm{MCTA} = Detection \times Tracking^{\mathrm{SCT}} \times Tracking^{\mathrm{ICT}} = SCTA \times Tracking^{\mathrm{ICT}}$
$= \left(\frac{2 \times Precision \times Recall}{Precision + Recall}\right)\left(1 - \frac{\sum_t mme_t^s}{\sum_t tp_t^s}\right)\left(1 - \frac{\sum_t mme_t^c}{\sum_t tp_t^c}\right)$

Tracking results

Sub-dataset	Evaluation metric	Comb1	Comb2	Comb3	Comb4	USC-Vision [1]	NLPR [2]	Hfutdspmct [3]	CRIPAC-MCT [4]
Dataset1	SCTA	0.6796				0.6448	0.6625	0.4301	0.1752
	$Tracking^{ICT}$	0.8851	0.8851	0.8665	0.8789	0.9288	0.6220	0.6534	0.7111
	MCTA	0.6015	0.6015	0.5889	0.5973	0.5989	0.4120	0.2810	0.1246
Dataset2	SCTA	0.7655				0.7358	0.6904	0.4598	0.1636
	$Tracking^{ICT}$	0.8842	0.8793	0.8818	0.8768	0.8691	0.6942	0.6122	0.7510
	MCTA	0.6769	0.6732	0.6751	0.6713	0.6260	0.4793	0.2815	0.1075
Dataset3	SCTA	0.6819				0.5476	0.6312	0.1475	0.0971
	$Tracking^{ICT}$	0.5461	0.5329	0.5329	0.5000	0.1014	0.2953	0.2432	0.1143
	MCTA	0.3724	0.3634	0.3634	0.3410	0.0555	0.1864	0.0359	0.0111
Dataset4	SCTA	0.8658				0.6262	0.6597	0.2064	0.0720
	$Tracking^{ICT}$	0.6270	0.6151	0.5992	0.6071	0.5437	0.4308	0.2944	0.2950
	MCTA	0.5429	0.5326	0.5188	0.5257	0.3404	0.2842	0.0608	0.0213
Average MCTA		0.5484	0.5427	0.5366	0.5338	0.4052	0.3405	0.1648	0.0661

Denotation	Feature combination	Denotation	Feature combination
Comb1	Holistic color, 2WGMMF, regional color/texture, couple	Comb3	Holistic color, 2WGMMF, couple
Comb2	2WGMMF, regional color/texture, couple	Comb4	Holistic color, regional color/texture, couple

References

[1] Y. Cai and G. Medioni, “Exploring context information for inter-camera multiple target tracking,” in Proc. IEEE WACV, 2014, pp. 761-768.
[2] L. Cao, W. Chen, X. Chen, S. Zheng, and K. Huang, “An equalized global graphical model-based approach for multi-camera object tracking,” arXiv:1502.03532v2, 2016.
[3] “Multi-Camera Object Tracking challenge,” [online] http://mct.idealtest.org/index.html.
[4] W. Chen, L. Cao, X. Chen, and K. Huang, “A novel solution for multi-camera object tracking,” in Proc. IEEE ICIP, 2014, pp. 2329-2333.