Simultaneous Multi-Person Tracking and Activity Recognition based on Cohesive Cluster Search

Wenbo Li, Yi Wei, Siwei Lyu, Ming-Ching Chang

Submission to CVIU

In the supplementary material, we provide the full notation of video activities, problem formulation and visual tracking (Table A.1); notations for instance, graph and hypergraph (Table A.2); component probabilities for the pairwise interaction activities (Table A.3); probabilistic formulation for the pairwise interactions (Table A.4).

Table A.1: *Notations for video activities, problem formulation and visual tracking.*

	Symbol	Description
Video Activities	$\mathtt{x}$	a target trajectory
	$\mathtt{a}$	an individual activity label, *e.g.* $\in \{standing, walking, running\}$
	$\mathtt{i}$	a pairwise interaction label, *e.g. approaching* (AP), *facing-each-other* (FE), *standing-in-a-row* (SR), ...
	$\mathtt{c}$	a collective activity label, *e.g.*, CROSSING, WALKING, GATHERING, ...
	b	number of observed targets (tracklets)
	T	video time window of length τ prior to time t, *i.e.*, $T = [t-\tau, t]$
	D	person detections (bounding boxes)
	X	target trajectories, $X_T = \{\mathtt{x}_1, \ldots, \mathtt{x}_b\}$
	A	individual activity classes, $A_T = \{\mathtt{a}_1, \ldots, \mathtt{a}_b\}$
	I	pairwise interaction classes, $I_T = \{\mathtt{i}_{1,2}, \mathtt{i}_{1,3}, \ldots, \mathtt{i}_{2,3}, \ldots, \mathtt{i}_{b-1,b}\}$
	C	collective activity classes, $C_T = \{\mathtt{c}_{t-\tau}, \ldots, \mathtt{c}_t\}$
	T', X', A', I', C'	existing entities prior to time window T, $X' = X_{T'}$, $A' = A_{T'}$, $I' = I_{T'}$, $C' = C_{T'}$
	n_A	number of individual activity classes, $n_A = 2$ in the CAD and Augmented-CAD datasets, $n_A = 3$ in the New-CAD dataset
	n_I	number of interaction classes, which is also the number of sub-hypergraphs used in our method, $n_I = 8$ in the CAD and Augmented-CAD datasets, $n_I = 9$ in the New-CAD dataset
	n_C	number of collective activity classes, $n_C = 5$ in CAD, $n_C = 6$ in Augmented-CAD, $n_C = 6$ in New-CAD datasets
Problem Formulation	$\mathbf{Pr}$	a joint distribution
	f_1, f_2, f_3	confidence terms from the decomposition of $\mathbf{Pr}$
	$\varphi_1, \varphi_2, \varphi_3$	clique potential functions in the Markov random field
	X^*, A^*, I^*, C^*	updated terms of X, A, I, C after an optimization stage, respectively
	$X^\ddagger, A^\ddagger$	updated terms of X^*, A^* after an optimization stage, respectively
	p_{ds}	the *distance* likelihood term for estimating the interaction between two targets
	p_{gc}	the *group connectivity* term for estimating the interaction between two targets
	p_{aa}	the *individual activity agreement* term for estimating the interaction between two targets
	p_{dc}	the *distance change type* likelihood term for estimating the interaction between two targets
	p_{dr}	the *facing direction* likelihood term for estimating the interaction between two targets
	p_{fs}	the *frontness/sideness* likelihood term for estimating the interaction between two targets
Tracking	$\bar{\mathtt{x}}$	a candidate tracklet
	$\bar{X}$	the set of all candidate tracklets
	$\bar{\mathtt{a}}$	a (putative) individual activity of a candidate tracklet
	$\bar{A}$	the set of (putative) individual activities for all candidate tracklets
	θ_a	the *appearance* similarity for tracklet linking
	τ_a	time threshold for appearance-based tracklet linking
	$\oplus$	operator $\oplus$ represents the association of two tracklets
	h	the number of hypothetical tracklets to generate from an existing tracklet $\mathtt{x}'_i$, $h = 9$

Table A.2: *Graph and hypergraph notations.*

	Symbol	Description
Hypergraph	$\mathcal{H}$	hypergraph $\mathcal{H} = (V, E, W)$
	$\mathcal{H}_{\mathcal{T}}$	tracking hypergraph $\mathcal{H}_{\mathcal{T}} = (V_{\mathcal{T}}, E_{\mathcal{T}}, W_{\mathcal{T}})$
	$\mathcal{H}_{\mathcal{R}}$	activity recognition hypergraph $\mathcal{H}_{\mathcal{R}} = (V_{\mathcal{R}}, E_{\mathcal{R}}, W_{\mathcal{R}})$
	V	the vertex set of a hypergraph
	E	the hyperedge set of a hypergraph
	W	the hyperedge weights of a hypergraph
	W_a	the *appearance* hyperedge weight, working with control parameter $\lambda_a = 30$
	W_d	the *facing-direction* hyperedge weight, working with control parameter $\lambda_d = 1$
	W_g	the *geometric similarity* hyperedge weight, working with control parameter $\lambda_g = 0.5$
	m	the hyperedge degree, *i.e.*, the number of incident vertices of the hyperedge
	$\mathbf{e}^m$	a m-degree hyperedge, $\mathbf{e}^m = \{v_1^{\mathbf{e}}, \ldots, v_m^{\mathbf{e}}\}$
	$\mathcal{C}$	a hyperedge cluster, which is a vertex set with interconnected hyperedges
	κ	number of vertes in a hypergraph cluster $\mathcal{C}$, $\kappa = \lvert\mathcal{C}\rvert$
	$E^{\mathcal{C}}$	the set of all incident hyperedges of a cluster $\mathcal{C}$
	Ψ	weighting function operated on a hypergraph cluster $\mathcal{C}$
	$\mathbf{y}$	the indicator vector to denote the vertex selection from $V \in \mathcal{H}$ to be included in $\mathcal{C}$
	ϵ	$\epsilon = \frac{1}{\kappa}$ used in weight normalization
	δ	$\delta_p = \frac{y_p}{\kappa}$ used in weight normalization
	$\mathbf{p}_{ij}$	image coordinate vector between two positions at i and j
	$\breve{\mathcal{H}}$	a sub-hypergraph indexed by β, *i.e.*, $\breve{\mathcal{H}}_\beta$
	$\breve{E}_\beta$	the hyperedges of the sub-hypergraph $\breve{\mathcal{H}}_\beta$ corresponding to the β-th interaction class
	$\breve{W}_\beta$	the hyperedge weights of the sub-hypergraph $\breve{\mathcal{H}}_\beta$ corresponding to the β-th interaction class
Graph	$\tilde{\mathcal{G}}$	graph $\tilde{\mathcal{G}} = (\tilde{V}, \tilde{E}, \tilde{W})$
	$\tilde{V}$	the vertex set of a graph; $\tilde{V}$ is associated with X' in this paper
	$\tilde{E}$	the edge set of a graph
	$\tilde{W}$	the edge weights of a graph
	e_{ij}	a graph edge connecting two vertices v_i and v_j
	p_{corr}	the correlation between the activities of two targets $\mathtt{x}_i$ and $\mathtt{x}_j$ used to calculate weight $\tilde{W}(e_{ij})$
	g	a function to calculate the correlation between the activities of two targets
	d	Eucludean distance between two targets in the image coordinate.
	ϕ_{ij}	the angle between the facing direction of $\mathtt{x}_i$ and the relative vector from $\mathtt{x}_i$ to $\mathtt{x}_j$.
	$\tilde{\mathcal{G}}_s$	sparse graph by discarding edges with small weights from $\tilde{\mathcal{G}}$
Indices	t, τ, f	video frame indices
	i, j, k, l	target tracklet indices
	p, q, r	hypergraph vertex indices
	α	the index for hypergraph clusters *e.g.* $\mathcal{C}_\alpha, \mathcal{C}_\alpha^{\mathcal{T}}$ from $\mathcal{H}_{\mathcal{T}}$
	β	the index for interaction classes *e.g.* I_β; β is also the index for sub-hypergraphs *e.g.* $\breve{\mathcal{H}}_\beta$
	c	the index for collective activity classes C

Table A.3: *Component probabilities for the pairwise interaction activities. The parameters used in these component probabilities,* e.g*, the means and standard deviations are calculated from the training dataset.*

Component	Probability
Distance	$p_{ds}(\textit{within-effective-range}\|x_i, x_j) = \delta(\|\frac{d_{ij}-\mu_{ds}}{\sigma_{ds}}\| \leq b), d_{ij} \sim \mathcal{N}(\mu_{ds}, \sigma_{ds})$, where $\mathcal{N}$ denotes normal distribution
Group connectivity	$p_{gc}(\texttt{GC}\|x_i, x_j)$, where $\texttt{GC} \in \{\textit{connect}, \textit{not-connect}\}$ $p_{gc}(\textit{connect}\|x_i, x_j) = g(d_{ij}, \phi_{ij}, \|\nu_i\|, \|\nu_j\|)$, where $g(\cdot)$ is defined in Eq.10 $p_{gc}(\textit{not-connect}\|x_i, x_j) = 1 - g(d_{ij}, \phi_{ij}, \|\nu_i\|, \|\nu_j\|)$
Individual activity agreement	$p_{aa}(a_i = \texttt{AA}_1, a_j = \texttt{AA}_2) = \sqrt{p(\texttt{AA}_1\|\nu_i) \cdot p(\texttt{AA}_2\|\nu_j)}$, where $\texttt{AA}_1, \texttt{AA}_2 \in \{\textit{standing}, \textit{walking}, \textit{running}\}$
Distance-change type	$p_{dc}(\texttt{DC}\|x_i, x_j)$, where $\texttt{DC} \in \{\textit{decreasing}, \textit{unchanging}, \textit{increasing}\}$ $p_{dc}(\textit{decreasing}\|x_i, x_j) = 1 - sigmoid(d^g_{ij}, \mu_{d2u}, \sigma_{d2u})$, where $d^g_{ij} = d_{ij} - d'_{ij}$ $p_{dc}(\textit{unchanging}\|x_i, x_j) = \lambda * sigmoid(d^g_{ij}, \mu_{d2u}, \sigma_{d2u}) + (1 - \lambda) * (1 - sigmoid(d^g_{ij}, \mu_{u2i}, \sigma_{u2i}))$; if $d^g_{ij} < \mu_{d2u}$, $\lambda = 1$; if $d^g_{ij} \geq \mu_{u2i}$, $\lambda = 0$; otherwise, $\lambda = 1 - \frac{d^g_{ij}-\mu_{d2u}}{\mu_{u2i}-\mu_{d2u}}$ $p_{dc}(\textit{increasing}\|x_i, x_j) = sigmoid(d^g_{ij}, \mu_{u2i}, \sigma_{u2i})$
Facing direction	$p_{dr}(\texttt{DR}\|x_i, x_j) \in \{0, 1\}$, where $\texttt{DR} \in \{\textit{same}, \textit{opposite}, \textit{frequent-changing}\}$
Frontness/sideness	$p_{fs}(\texttt{FS}\|x_i, x_j)$, where $\texttt{FS} \in \{\textit{frontness}, \textit{sideness}\}$ $p_{fs}(\textit{frontness}\|x_i, x_j) = \max(cos(\mathbf{p}_{ij}, x_i), cos(\mathbf{p}_{ij}, x_j))$ $p_{fs}(\textit{sideness}\|x_i, x_j) = 1 - p_{fs}(\textit{frontness}\|x_i, x_j)$

Table A.4: *Probabilistic formulations for the pairwise interactions $p(\texttt{i}_{ij} = \beta)$. We define* dancing-together *(DT) as a new interaction activity class to deal with the new collective activity "dancing" in the Augmented-CAD.*

Pairwise Interaction $p(\texttt{i}_{ij} = \beta)$	Associated Collective Activity (C)	Probabilistic Formulation
facing-each-other (β =FE)	TALKING	$p_{ds}(\textit{within-effective-range}) \cdot p_{gc}(\textit{connect}) \cdot p_{aa}(\textit{standing}, \textit{standing}) \cdot$ $p_{dc}(\textit{unchanging}) \cdot p_{dr}(\textit{opposite}) \cdot p_{fs}(\textit{frontness})$
standing-in-a-row (β =SR)	QUEUING	$p_{ds}(\textit{within-effective-range}) \cdot p_{gc}(\textit{connect}) \cdot p_{aa}(\textit{standing}, \textit{standing}) \cdot$ $p_{dc}(\textit{unchanging}) \cdot p_{dr}(\textit{same}) \cdot p_{fs}(\textit{frontness})$
standing-side-by-side (β =SS)	WAITING	$p_{ds}(\textit{within-effective-range}) \cdot p_{gc}(\textit{connect}) \cdot p_{aa}(\textit{standing}, \textit{standing}) \cdot$ $p_{dc}(\textit{unchanging}) \cdot p_{dr}(\textit{same}) \cdot p_{fs}(\textit{sideness})$
dancing-together (β =DT)	DANCING	$p_{ds}(\textit{within-effective-range}) \cdot p_{gc}(\textit{connect}) \cdot p_{aa}(\textit{walking}, \textit{walking}) \cdot$ $p_{dc}(\textit{unchanging}) \cdot p_{dr}(\textit{frequent-chaning}) \cdot p_{fs}(\textit{sideness})$
approaching (β =AP)	GATHERING	$p_{ds}(\textit{within-effective-range}) \cdot p_{gc}(\textit{not-connect}) \cdot p_{aa}(\textit{walking}, \textit{walking}) \cdot$ $p_{dc}(\textit{decreasing}) \cdot p_{dr}(\textit{opposite}) \cdot p_{fs}(\textit{frontness})$
*walking-in-opposite-*directions (β =WO), *leaving* (β =LV)	DISMISSAL	$p_{ds}(\textit{within-effective-range}) \cdot p_{gc}(\textit{not-connect}) \cdot p_{aa}(\textit{walking}, \textit{walking}) \cdot$ $p_{dc}(\textit{increasing}) \cdot p_{dr}(\textit{opposite}) \cdot p_{fs}(\textit{frontness})$
walking-side-by-side (β =WS)	CROSSING WALKING (TOGETHER)	$p_{ds}(\textit{within-effective-range}) \cdot p_{gc}(\textit{connect}) \cdot p_{aa}(\textit{walking}, \textit{walking}) \cdot$ $p_{dc}(\textit{unchanging}) \cdot p_{dr}(\textit{same}) \cdot p_{fs}(\textit{sideness})$
running-side-by-side (β =RS)	JOGGING	$p_{ds}(\textit{within-effective-range}) \cdot p_{gc}(\textit{connect}) \cdot p_{aa}(\textit{running}, \textit{running}) \cdot$ $p_{dc}(\textit{unchanging}) \cdot p_{dr}(\textit{same}) \cdot p_{fs}(\textit{sideness})$
running-one-after-the-other (β =RR)	CHASING	$p_{ds}(\textit{within-effective-range}) \cdot p_{gc}(\textit{connect}) \cdot p_{aa}(\textit{running}, \textit{running}) \cdot$ $p_{dc}(\textit{unchanging}) \cdot p_{dr}(\textit{same}) \cdot p_{fs}(\textit{frontness})$