SANTIAGO, CHILE DECEMBER 11-18, 2015 ICCV15

Unsupervised Synchrony Discovery in Human Interaction

Wen-Sheng Chu[1], Jiabei Zeng[2], Fernando De la Torre[1], Jeffrey F. Cohn[1,3], and Daniel Messinger[4]

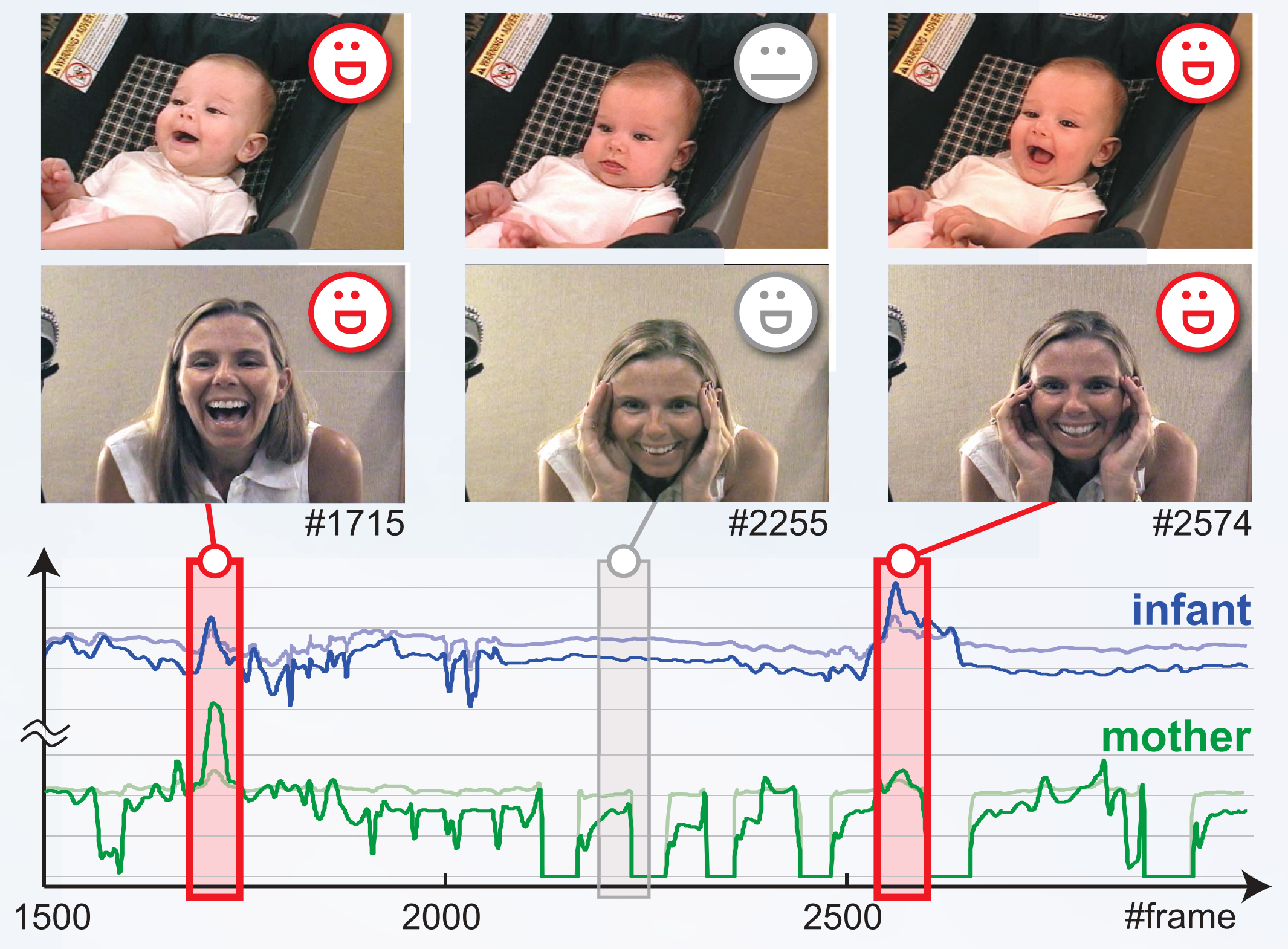
NEW

Problem

● **Unsupervised synchrony discovery (USD)**

△**Goal:** Discover interpersonal synchrony without supervision

△**Synchrony:** Defined as matched states between two or more persons



● Problem formulation

△For a pair of videos:

$$\max_{\{b_1,e_1,b_2,e_2\}} \quad f(\phi_{\mathsf{S}^1[b_1,e_1]}, \phi_{\mathsf{S}^2[b_2,e_2]}),$$

$$\text{subject to} \quad \ell \le e_i - b_i, \forall i \in \{1,2\}, |b_1 - b_2| \le T$$

△For a set of N videos:

$$\max_{\{b_i,e_i\}_{i=1}^N} \quad F\left(\{\phi_{\mathsf{S}^i[b^i,e^i]}\}_{i=1}^N\right)$$

$$\text{subject to} \quad \ell \le e_i - b_i, \max(|b_i - b_j|) \le T, \forall i \ne j,$$

where $F\left(\{\phi_{\mathsf{S}^i[b^i,e^i]}\}_{i=1}^N\right) = \sum_{i \ne j} f(\phi_{\mathsf{S}^i[b_i,e_i]}, \phi_{\mathsf{S}^j[b_j,e_j]})$.

● USD is challenging

△Non-convex and non-differentiable

△An exhaustive search costs $\mathcal{O}(n^4)$, which is computationally prohibitive for long videos.

● Departure from previous work

△**USD** vs **ESS** [1] / **STBB** [2]

- Temporal domain vs spatial domain
- Unsupervised learning vs supervised learning

△**USD** vs **ACA** [3]

- Synchrony discovery vs temporal clustering

△**USD** vs **TCD** [4]

- Specific search vs general search
- Discovery between two or more sequences
- New bounding functions and speed-up strategies

[1] C. H. Lampert, et al. Efficient subwindow search: A branch and bound framework for object localization, TPAMI, pp. 2129–214

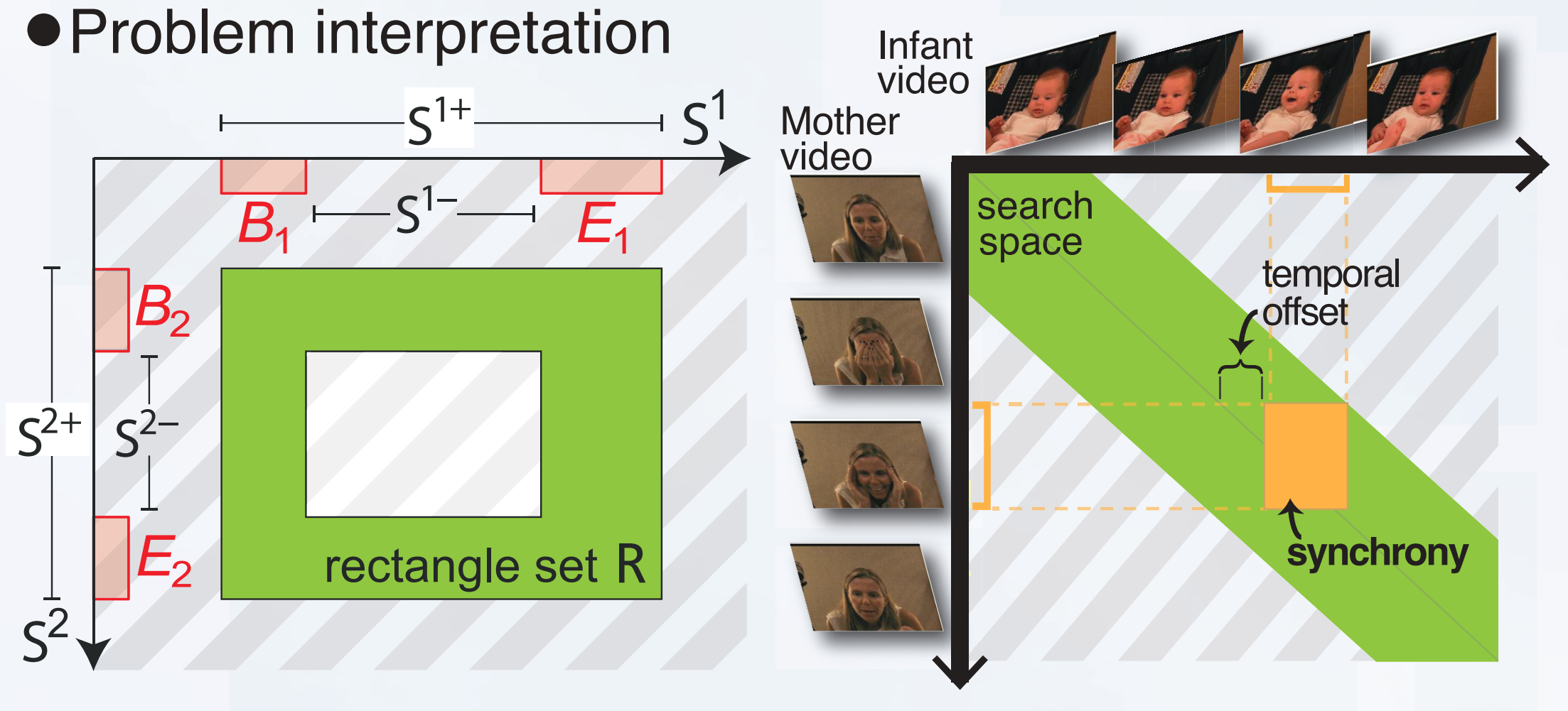
[2] J. Yuan, et al. Discriminative video pattern search for efficient action detection, TPAMI, 33:1728–1743, 2011.

[3] F. Zhou, et al. Unsupervised discovery of facial events" in CVPR, 2010.

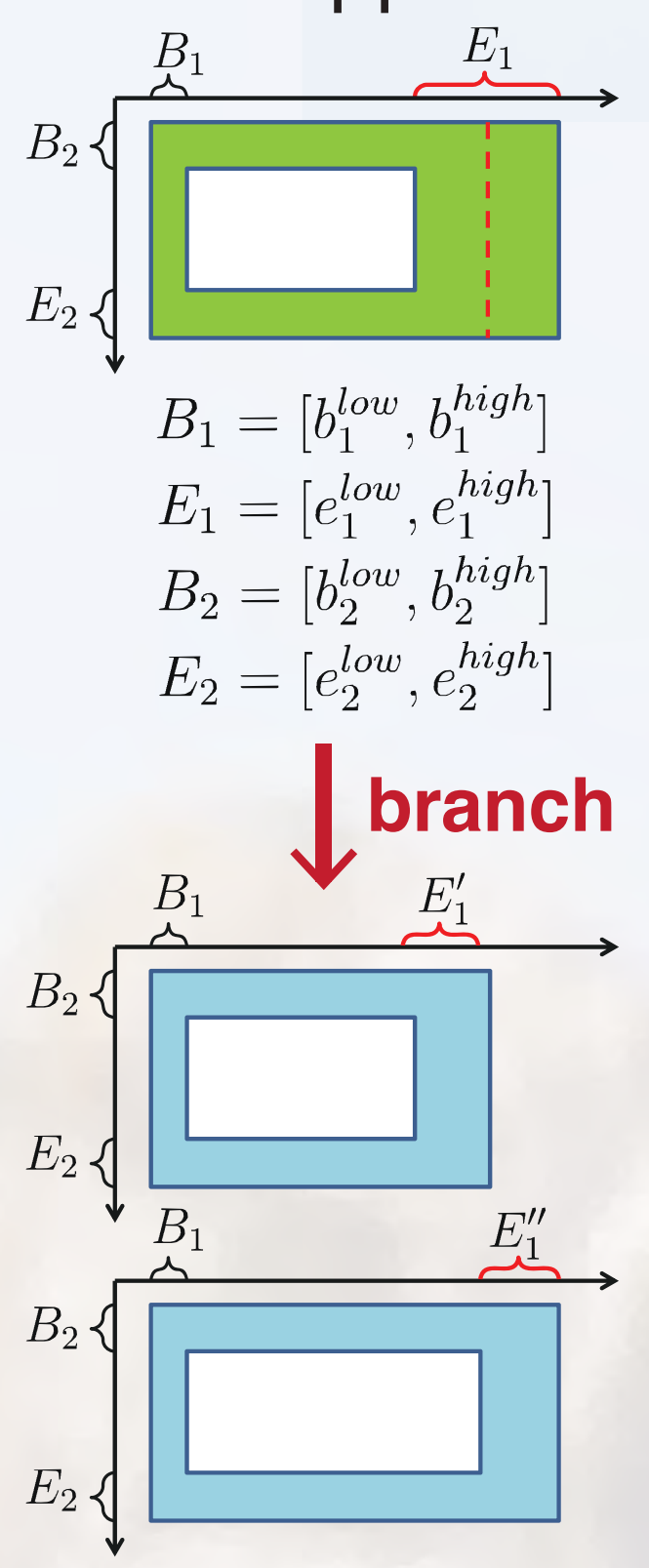
[4] W.-S. Chu, et al. Unsupervised temporal commonality discovery, ECCV 2012.

Efficient B&B Search

● Problem interpretation



● USD approach to **branch**-and-**bound**



B_1
E_1
B_2
E_2
$B_1 = [b_1^{low}, b_1^{high}]$
$E_1 = [e_1^{low}, e_1^{high}]$
$B_2 = [b_2^{low}, b_2^{high}]$
$E_2 = [e_2^{low}, e_2^{high}]$
branch
E_1'
E_1''

Algorithm 1: Unsupervised Synchrony Discovery

```
input : A synchronized video pair A, B; minimal
        discovery length $\ell$; commonality period $T$
output: Optimal intervals $\mathbf{r}^\star = [b_1, e_1, b_2, e_2]$
$L \leftarrow T + \ell$; // The largest possible searching period
Q ← empty priority queue; // Initialize Q
for $t \leftarrow 1$ to $(n-T-L+1)$ do
    R ← $[t, t+T] \times [t+\ell-1, t+T+L-1] \times [t-T, t+T] \times [t-T+\ell-1, t+T+L-1]$;
    Q.push(bound(R), R); // Fill in Q
end
R ← Q.pop(); // Initialize R
while |R| ≠ 1 do
    R → R_1 ∪ R_2;              // Split into 2 disjoint sets
    Q.push(bound(R_1), R_1);    // Push R_1 and its bound
    Q.push(bound(R_2), R_2);    // Push R_2 and its bound
    R ← Q.pop();                // Pop top state from Q
end
r* ← rect(R); // Retrieve the optimal rectangle
```

● New **bounding** functions (l_1, l_2, intersection, X^2 in [4])

△Cosine similarity

$$l_C(\mathtt{R}) = \frac{\sum_k h_k^{i-} h_k^{j-}}{\|\mathsf{S}^{i+}\|\|\mathsf{S}^{j+}\|} \le C(\mathbf{h}^i, \mathbf{h}^j) \le \frac{\sum_k h_k^{i+} h_k^{j+}}{\|\mathsf{S}^{i-}\|\|\mathsf{S}^{j-}\|} = u_C(\mathtt{R})$$

△Symmetrized KL divergence

$$l_D(\mathtt{R}) = \sum_k (\underline{h_k^i} - \overline{h_k^j})_+ (\ln \underline{h_k^i} - \ln \overline{h_k^j})_+$$

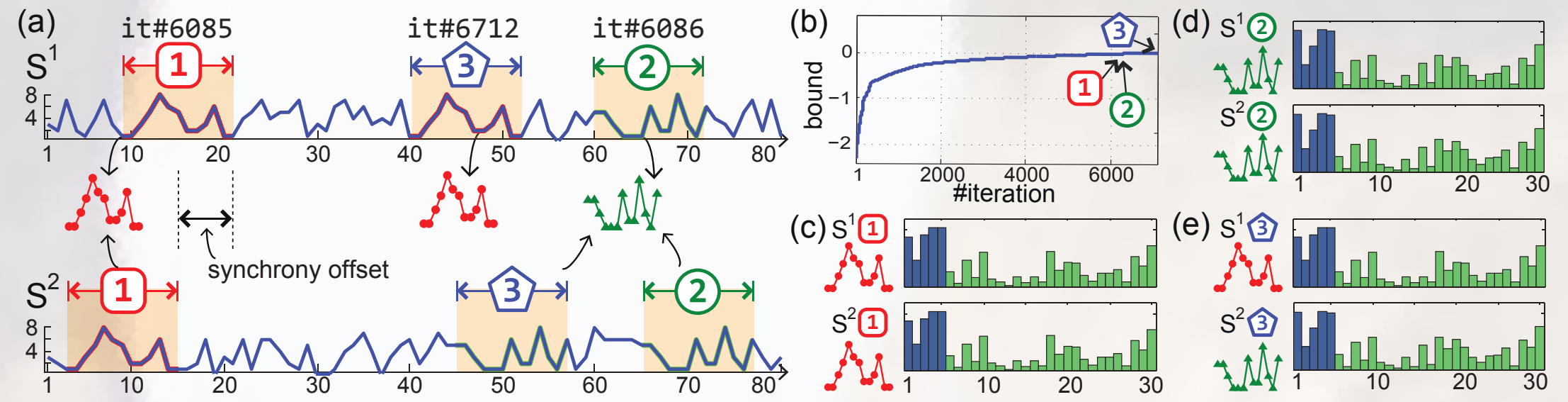
$$\le D(\mathbf{h}^i, \mathbf{h}^j) \le \sum_k (\overline{h_k^i} - \underline{h_k^j})(\ln \overline{h_k^i} - \ln \underline{h_k^j}) = u_D(\mathtt{R})$$

△Symmetrized cross entropy

$$l_E(\mathtt{R}) = \sum_k \left(-\underline{h_k^i} \log \overline{h_k^j} - \underline{h_k^j} \log \overline{h_k^i}\right)$$

$$\le E(\mathbf{h}^i, \mathbf{h}^j) \le \sum_k \left(-\overline{h_k^i} \log \underline{h_k^j} - \overline{h_k^j} \log \underline{h_k^i}\right) = u_E(\mathtt{R})$$

● Toy example



Extensions of USD

● By the nature of B&B algorithm, we extend USD to:

△Discover mMultiple synchronies

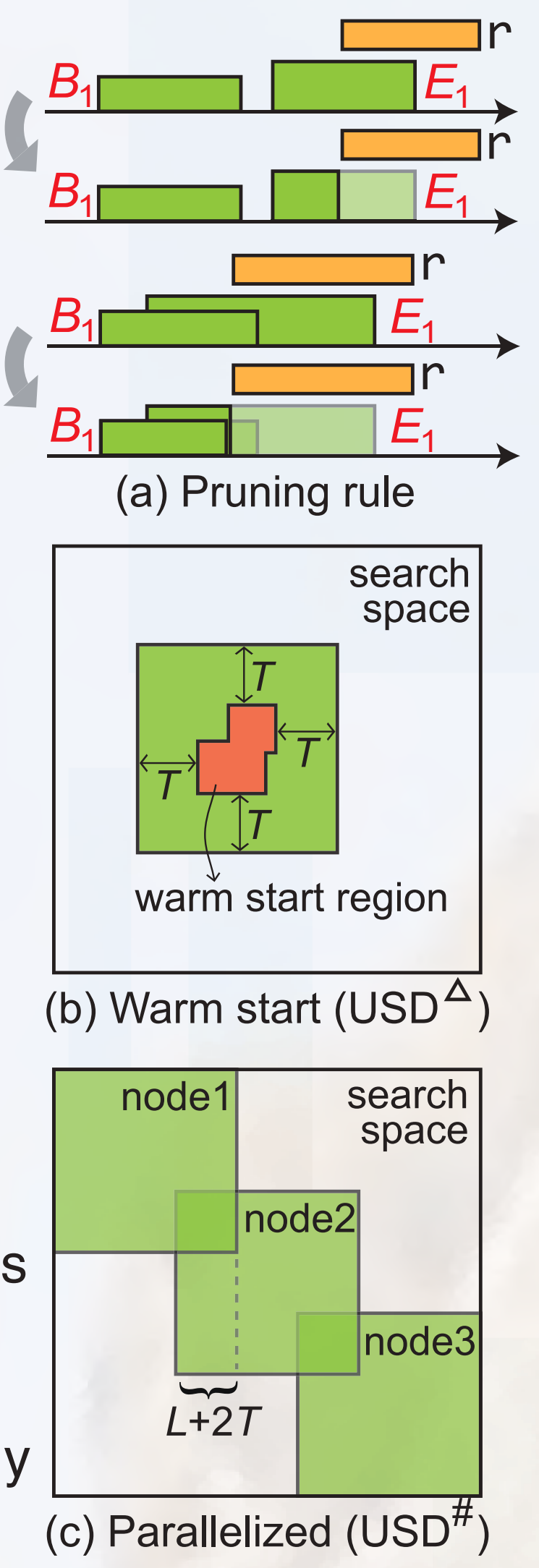
- Repeat USD algorithm multiple times
- **Strategy:** Safely discard undesired braches before starting the next USD.
- **In practice:** This strategy dramatically reduce the search space. In the toy example, search space reduced by 19% for the 2nd USD, and 25% for the 3rd.

△Warm start

- B&B identifies a solution quickly when neighborhood contains a clear optimum.
- **Strategy:** Estimate an initial solution with high quality (warm start region)
- **In practice:** Reduce computational cost to only few % of total iterations, and thus prune branches in main USD algorithm.

△Parallelized algorithm

- Speedup B&B with parallelism
- **Strategy:** Divide search into subproblems and perofrm discovery to each.
- **In practice:** The diagonal nature of USD gaurantees a global solution and an easily programmable and efficient algorithm.



(a) Pruning rule

(b) Warm start (USD$^\triangle$)

(c) Parallelized (USD$^\#$)

Experiment Settings

● **Datasets**

△Posed actions [5]

- 24 categories of actions, eg, jump, kick, run, walk, etc.
- 14 annotated sequences, 800~1600 frames each.

△Parent-infant interaction [6]

- 6 parent-infant video pairs, 3-minute each.
- Labels are smile, cry, neutral, occlusion.

△Social interaction [7]

- 48 participants in three-person groups.
- 2-minute videos coded with AUs (10,12,14,15,17,23,24).

● **Evaluation**

△Comparison via distance [4]

- Eg, l_1, l_2, intersection, X^2, KL-divergence, cross entropy, etc.

△Comparison via expert labels using recurrence quality [8]

$$\mathcal{Q}(\mathbf{r}) = \frac{1}{C \prod_i n_i} \sum_c \sum_{(i,j) \in \mathcal{A}} \sum_{p,q} I(\mathbf{Y}_i^c[p] = \mathbf{Y}_j^c[q])$$

[5] CMU Mocap. http://mocap.cs.cmu.edu/.

[6] D. S. Messinger, et al. Automated measurement of facial expression in infant–mother interaction. Infancy, 14(3):285–305, 2009.

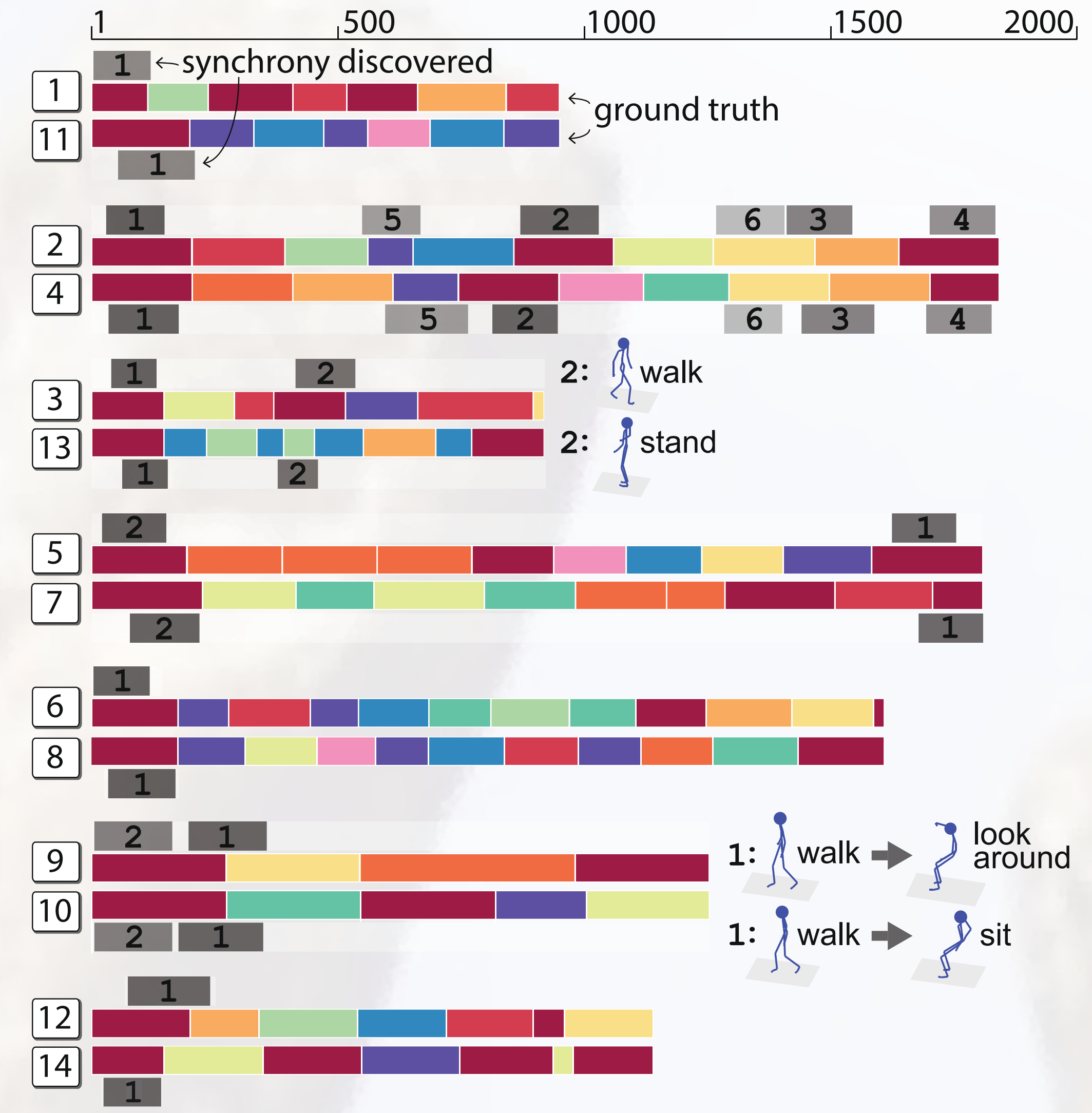
[7] J. M. Girard, et al. Spontaneous facial expression in unscripted social interactions can be measured automatically. Behavior Research Methods, Advance online publication, 2015.

[8] E. Delaherche, et al. Interpersonal synchrony: A survey of evaluation methods across disciplines. TAFFC, 3(3):349–365, 2012.

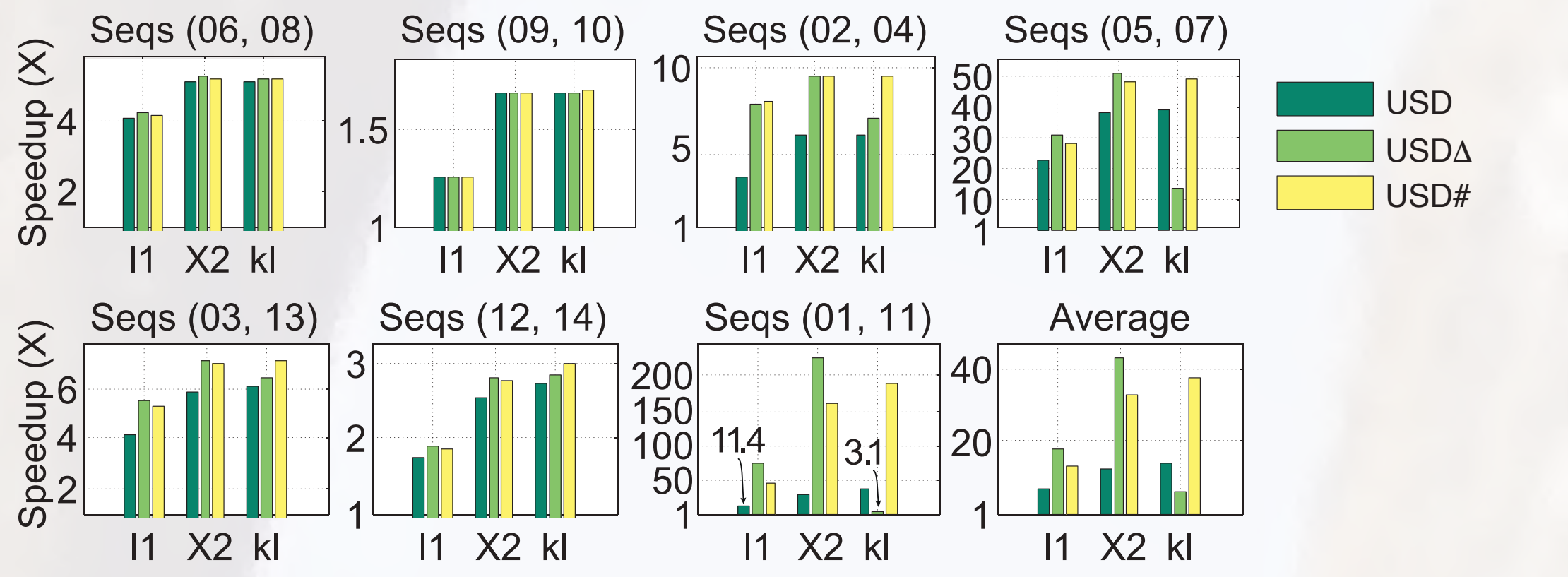
USD for Posed Actions

● **Posed actions** [5]

△Within-person synchrony for multiple actions from subject #86.



△Speedup evaluation against exhaustive sliding window (SW)

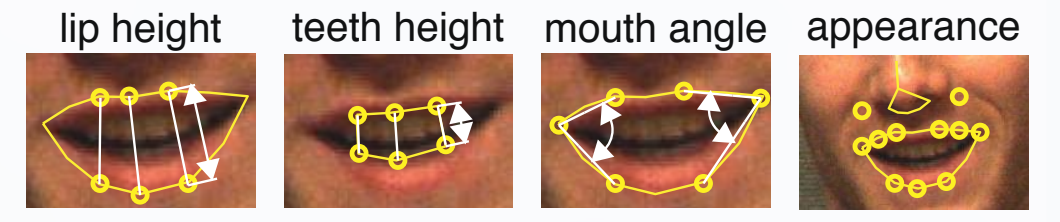


△Distance and quality analysis on all 7 pairs

	Pair	(1,11)	(2,4)	(3,13)	(5,7)	(6,8)	(9,10)	(12,14)	Avg.
χ^2-distance	USD	**6.3**	**1.2**	**4.7**	**2.6**	**0.1**	**0.2**	**11.9**	**3.9**
	$\mathrm{SW}_5^\star$	6.5	1.3	6.7	5.4	**0.1**	0.4	12.0	4.6
	$\mathrm{SW}_{10}^\star$	6.7	2.7	6.7	10.1	0.2	0.7	14.3	5.9
	SW_5^μ	97.1	76.9	81.4	64.2	89.3	172.0	334.5	130.8
	SW_5^σ	33.8	74.4	53.8	28.2	79.2	117.7	345.1	104.6
	SW_{10}^μ	94.8	77.3	81.8	63.2	87.1	170.2	327.2	128.8
	SW_{10}^σ	34.3	74.1	54.2	28.3	79.4	117.8	341.5	104.2
Rec. consistency	USD	0.89	**0.85**	0.46	**0.90**	**1.00**	0.64	**0.76**	**0.79**
	$\mathrm{SW}_5^\star$	**0.95**	0.81	**0.50**	0.84	**1.00**	**0.69**	0.73	**0.79**
	$\mathrm{SW}_{10}^\star$	**0.95**	0.75	**0.50**	0.64	**1.00**	0.55	0.00	0.63
	SW_5^μ	0.07	0.32	0.09	0.07	0.08	0.13	0.12	0.12
	SW_5^σ	0.16	0.33	0.25	0.20	0.21	0.29	0.22	0.24
	SW_{10}^μ	0.08	0.31	0.09	0.07	0.09	0.13	0.12	0.13
	SW_{10}^σ	0.19	0.33	0.26	0.21	0.22	0.29	0.23	0.25

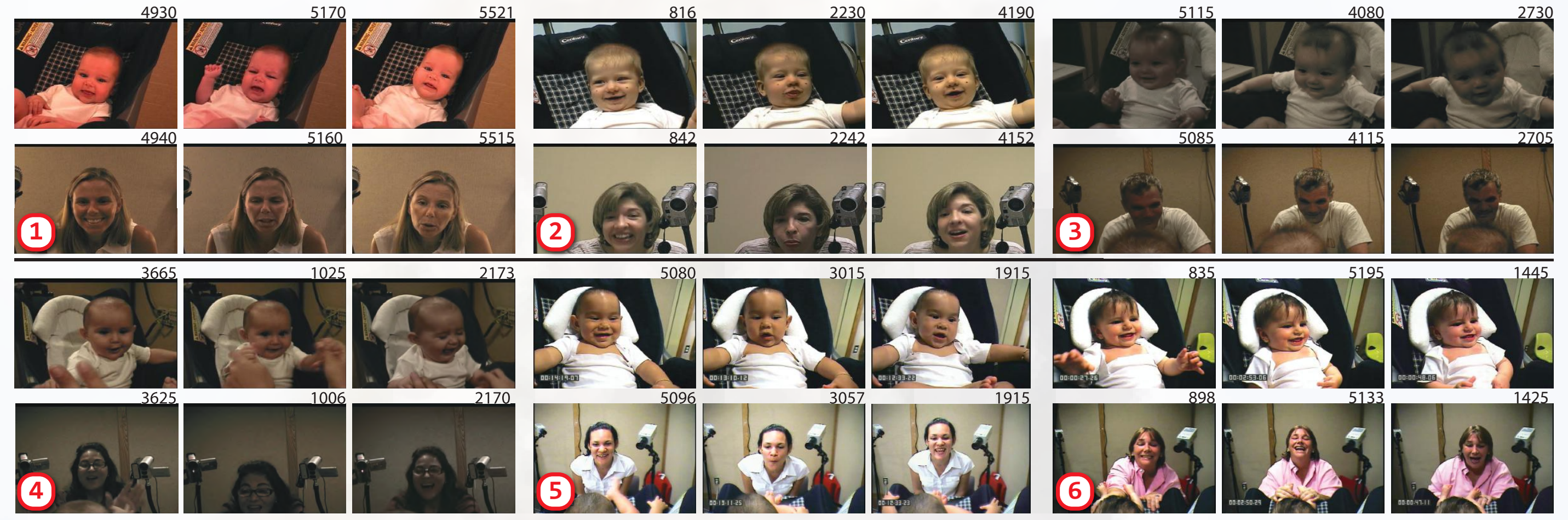
USD for Spontaneous Social Interactions

● Synchrony in **parent-infant interaction** [6]



△Critical for children in early social development.

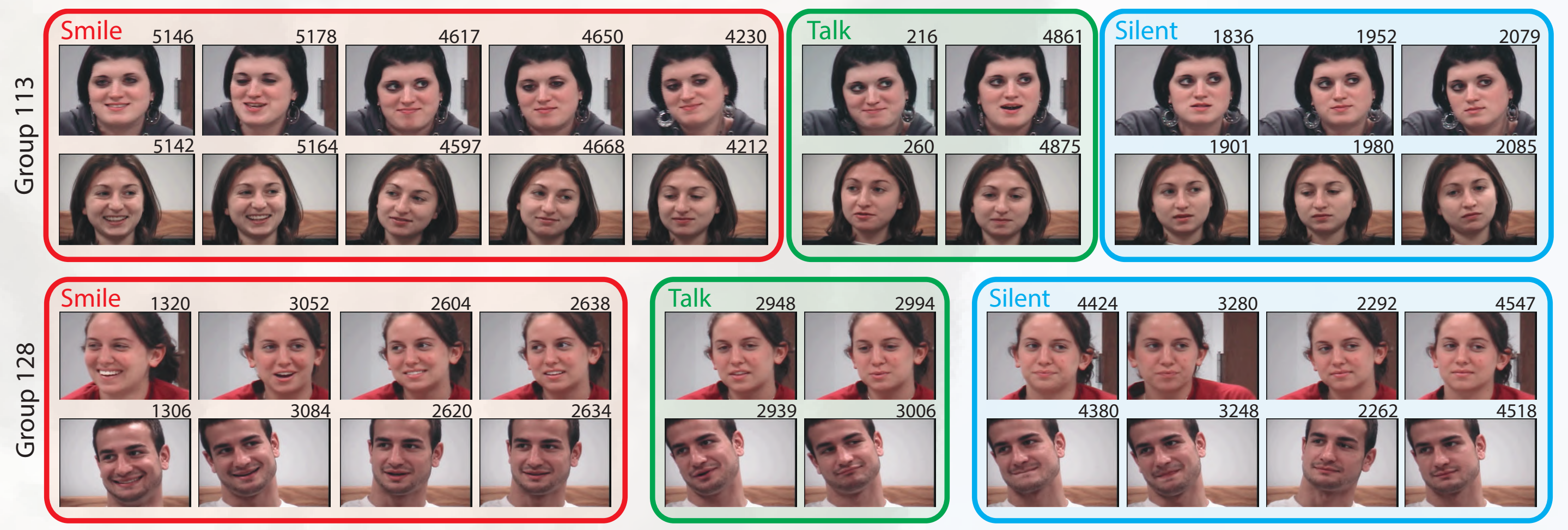
△Most synchony, represented in shape, was discovered as co-ocurring smiles.



● Synchrony in **dyadic social interaction** [7]

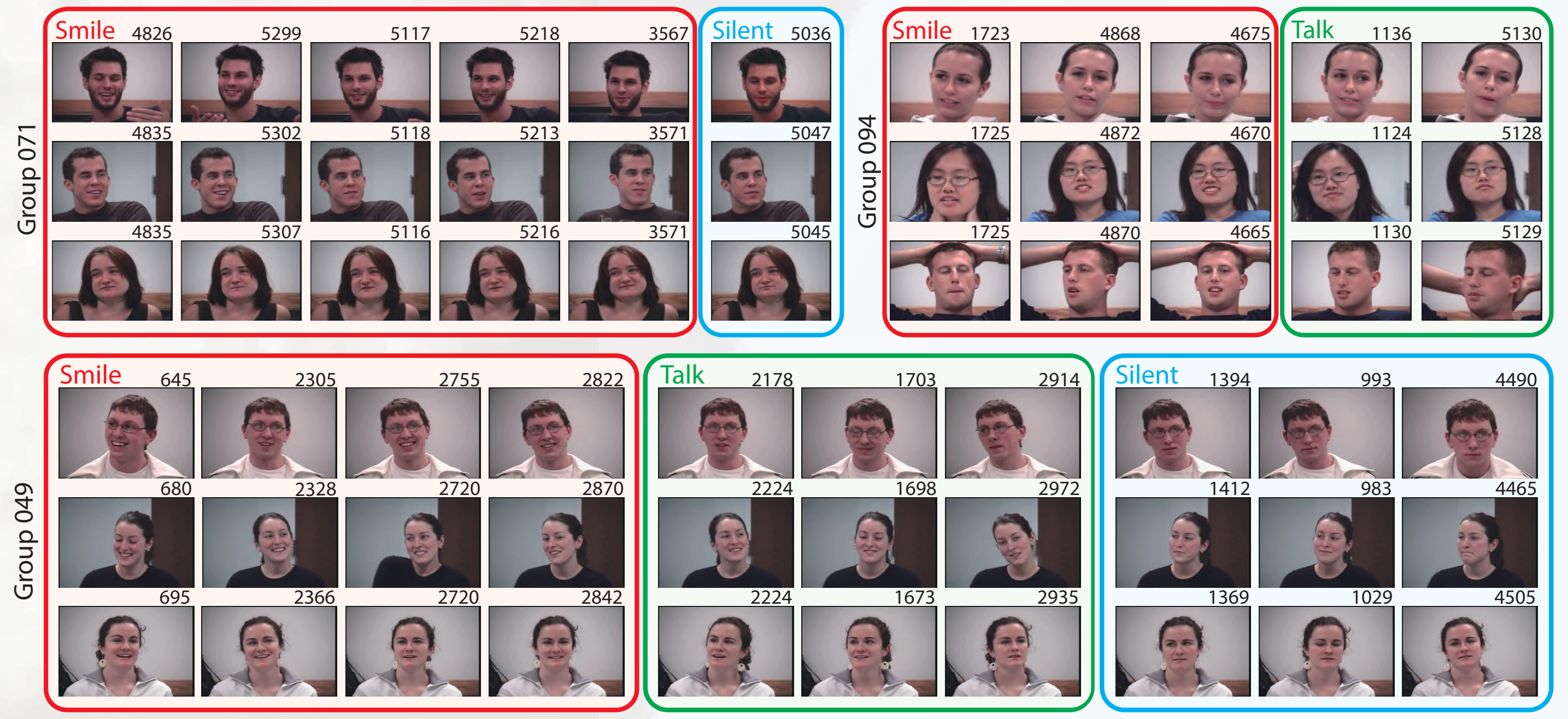
△Represent each face by concatenating appearance (SIFT) and shape (landmarks).

△Most synchony was discovered as co-ocurring smiles, talking and slience.



● Synchrony in **triadic social interaction** [7]

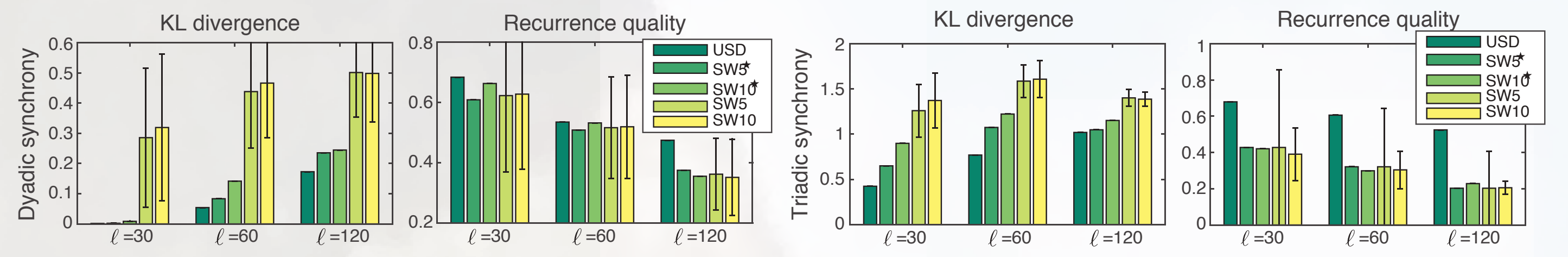
△Most synchony was discovered as co-ocurring smiles, talking and slience.



● **Quantative analysis**

△Compared USD with exhaustive sliding window (SW) with step sizes 5 and 10.

△Evaluated dyadic and triadic discovery in KL divergence and recurrence quality.


KL divergence
Recurrence quality
Dyadic synchrony
Triadic synchrony
USD
SW5*
SW10*
SW5
SW10
ℓ =30
ℓ =60
ℓ =120