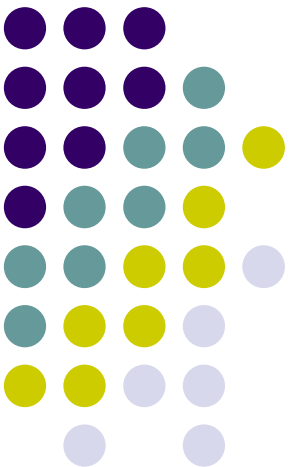
Efficient Content-Based Retrieval of Motion Capture Data

Rahul Singhai
Behjat Siddiqui
Sachin Patil



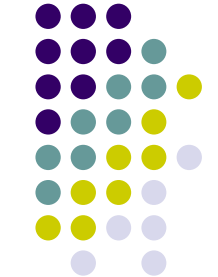
Problem

- Identify and extract those motion clips from a large motion database that are in some sense “logically similar” to a query motion (and user specification)

Crucial Notion -Similarity

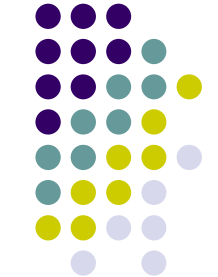
- Two motion clips "logically similar" if they are variation (spatial or temporal) of same action
- Logical Similarity ~Spatio-temporal invariance
- Logical Similarity != Numerical Similarity

Approach

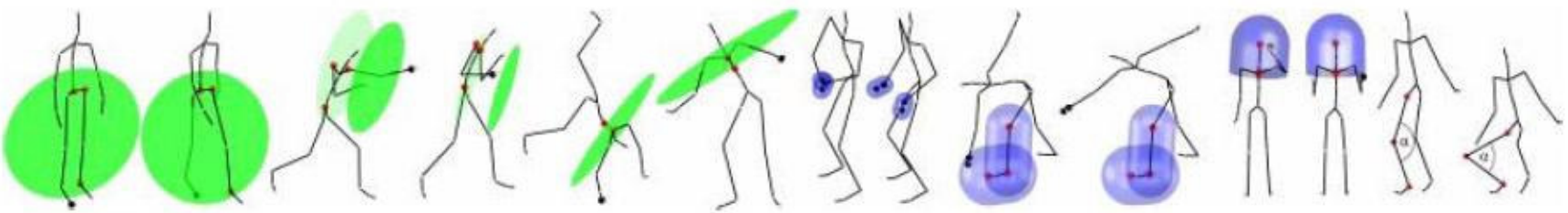


- Geometric Features
  - Describe motion at high semantic level
  - Robust against spatial variations
- Adaptive Segmentation
  - Robust against temporal variations
  - Transforms motion into sequence of geometric configurations

Geometric Features

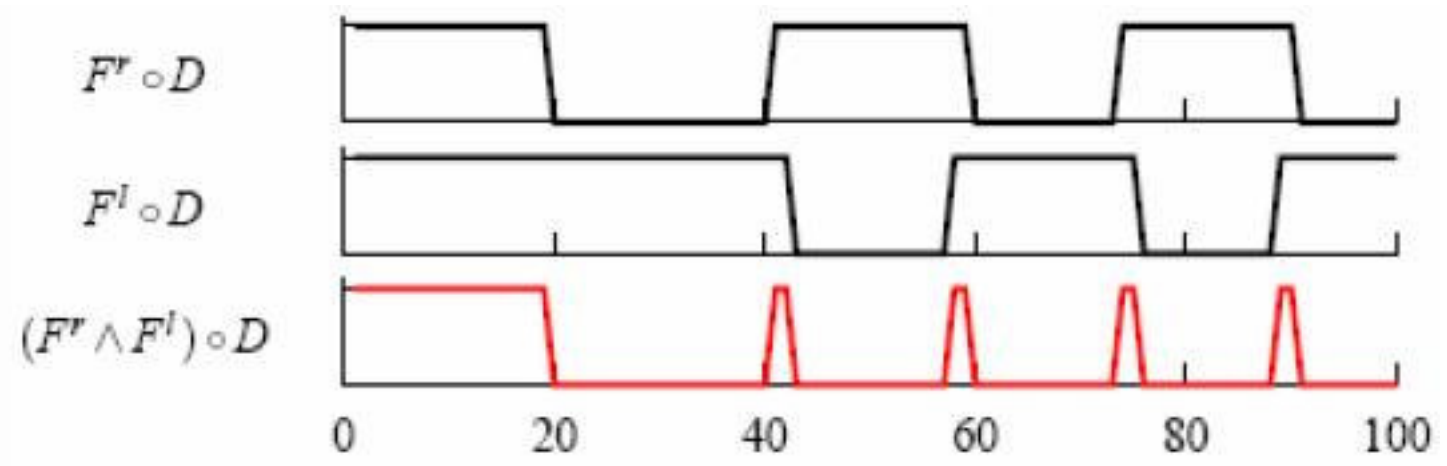


- Boolean Functions
- Describe relation between body parts

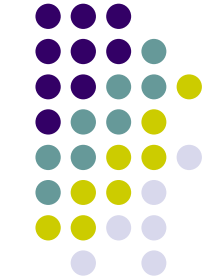


Geometric Features

- Concatenate to form feature vector F(P)
- Invariant to Euclidean motion, scaling, spatial variations
- Coarse description of motion



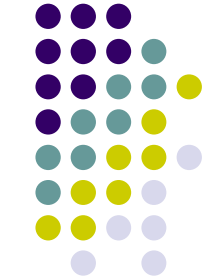
Geometric Features



- Total of 31 features
- Divide pose-dependant space into “octants” defined by three intersecting planes (above/below, left/right, front/behind)

Feature set	Features and corresponding abbreviations
F_ℓ	right/left foot in front (F_ℓ^1/F_ℓ^2), right/left foot raised (F_ℓ^3/F_ℓ^4), right/left foot fast (F_ℓ^5/F_ℓ^6), right/left knee bent (F_ℓ^7/F_ℓ^8), right/left leg sideways (F_ℓ^9/F_ℓ^{10}), legs crossed (F_ℓ^{11})
F_u	right/left hand in front ($F_\mathrm{u}^1/F_\mathrm{u}^2$), right/left hand raised ($F_\mathrm{u}^3/F_\mathrm{u}^4$), right/left arm sideways ($F_\mathrm{u}^5/F_\mathrm{u}^6$), right/left elbow bent ($F_\mathrm{u}^7/F_\mathrm{u}^8$), right/left hand fast ($F_\mathrm{u}^9/F_\mathrm{u}^{10}$), arms crossed ($F_\mathrm{u}^{11}$), hands touching ($F_\mathrm{u}^{12}$)
F_m	right/left hand touching any leg ($F_\mathrm{m}^1/F_\mathrm{m}^2$), right/left hand touching head or neck ($F_\mathrm{m}^3/F_\mathrm{m}^4$), right/left hand touching hip area ($F_\mathrm{m}^5/F_\mathrm{m}^6$), torso bent ($F_\mathrm{m}^7$), root fast ($F_\mathrm{m}^8$)

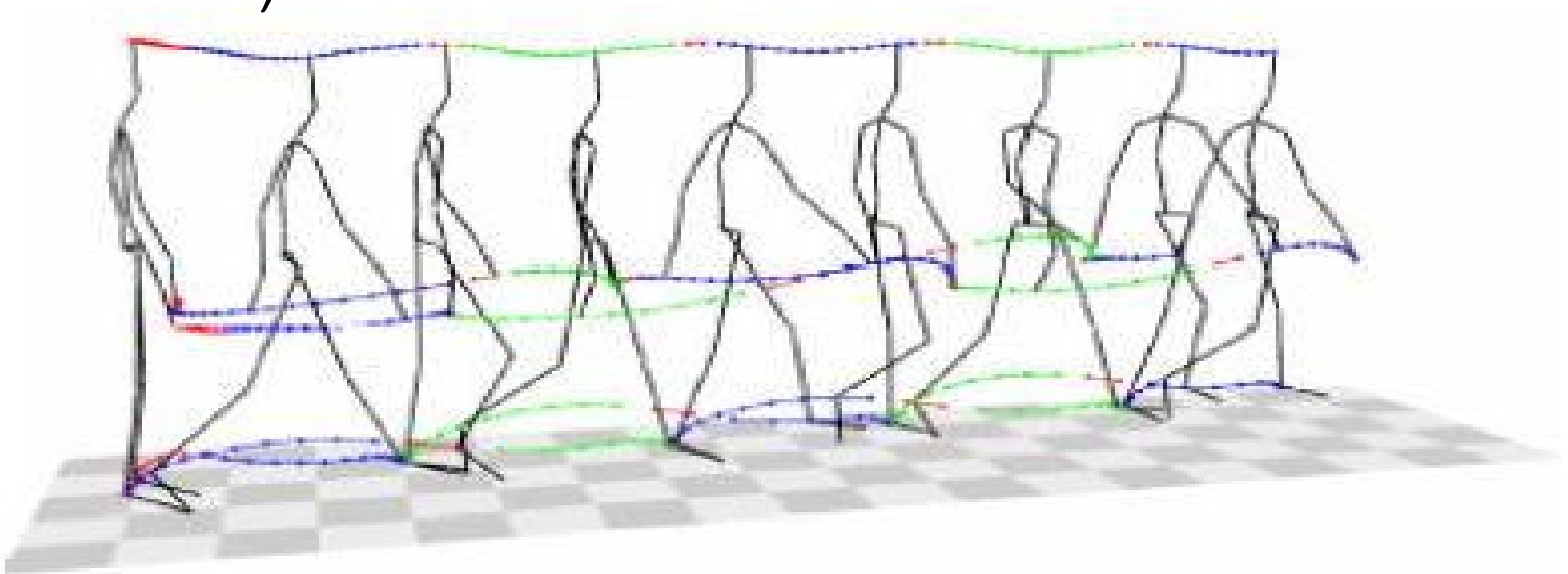
Geometric Features



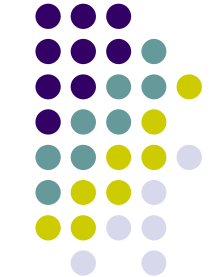
- Advantages
  - Logical similar motions may exhibit considerable spatial as well as temporal deviations. Pose-based geometric features are invariant under such spatial variations
- Disadvantages
  - How many to select
    - Over or under specification
  - How to select
  - Broad range of motions

Adaptive Segmentation

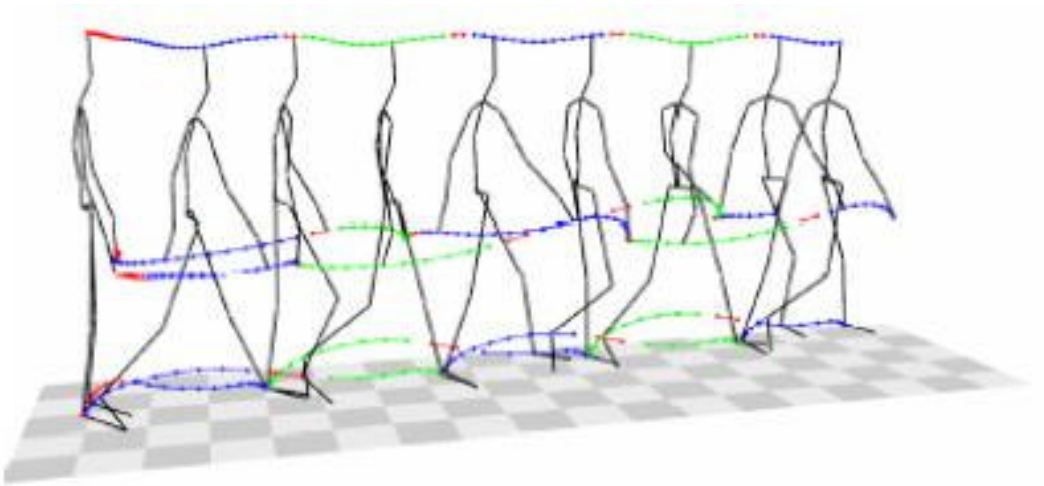
- Two poses F-equivalent if F(P1) = F(P2).
- F-run of D is subsequence of D consisting of consecutive F-equivalent poses.
- F-segment is F-run of maximal length (10 in fig. below)



Adaptive Segmentation



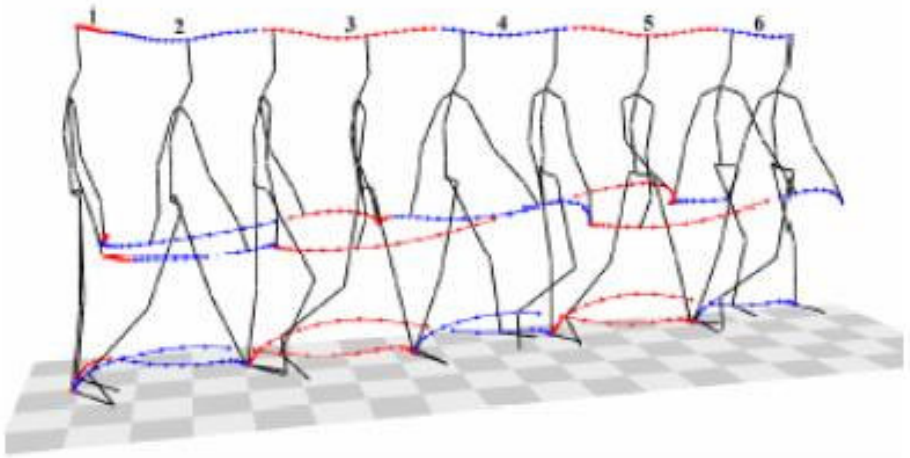
- Each segment correspond to unique feature vector.
- Segments induce feature vector sequence F[D].

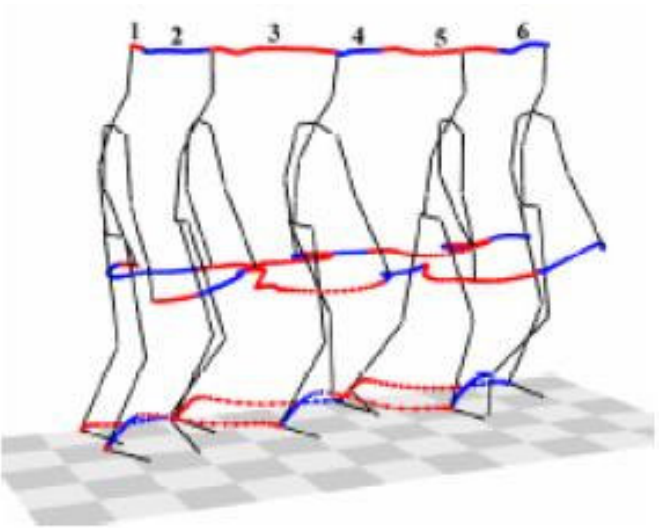


$$F^2[D_{\text{walk}}] = \left(\binom{1}{1}, \binom{0}{1}, \binom{1}{1}, \binom{1}{0}, \binom{1}{1}, \binom{0}{1}, \binom{1}{1}, \binom{1}{0}, \binom{1}{1}, \binom{0}{1}\right).$$

Adaptive Segmentation

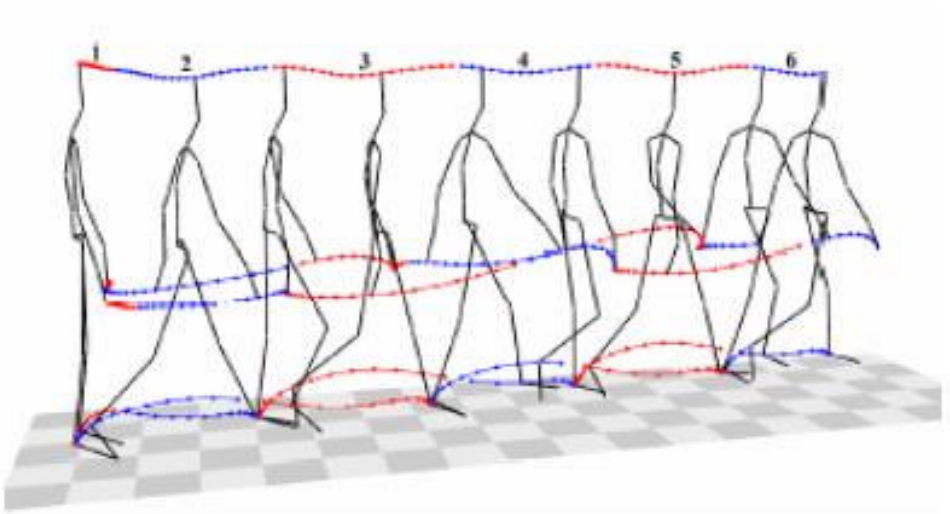
- **Time invariance** : Two motions that differ by some deformation of the time axis will yield the same F-feature sequences.

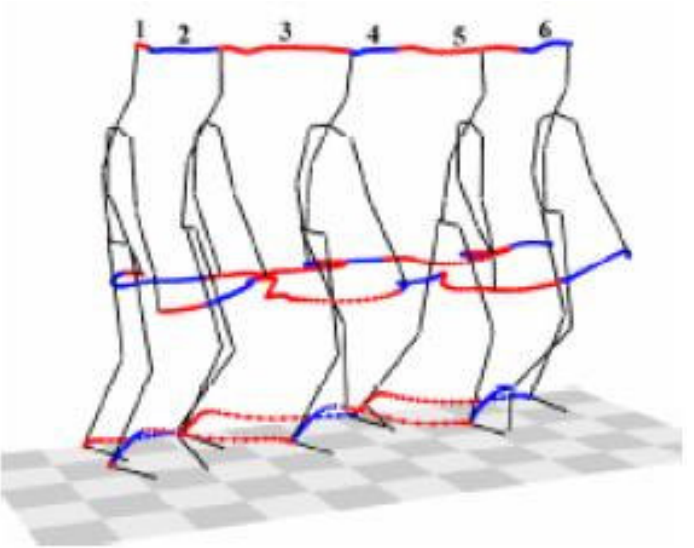




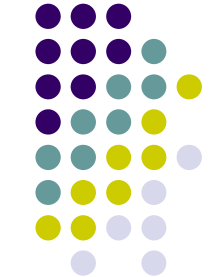
Adaptive Segmentation

- Adapts to selected features
  - Fine features = Many short segments
  - Coarse features = Lesser long segments





Adaptive Segmentation



- Compare at “segment level” rather than “frame level”
  - Use F-feature sequence F[D]
  - Computational saving
  - Spatio-temporal invariance incorporated

Indexing and Retrieval

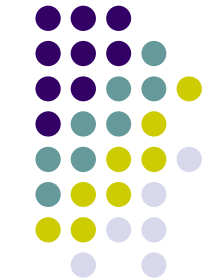
- Fix feature set
- Construct query-independent index structure of F – inverted list

$$L(\binom{1}{1}) = \{1,3,5,7,9\},\ L(\binom{0}{1}) = \{2,6,10\},\ L(\binom{1}{0}) = \{4,8\},\ L(\binom{0}{0}) = \emptyset.$$

$$F[Q] = (\binom{1}{0},\binom{1}{1},\binom{0}{1}).$$

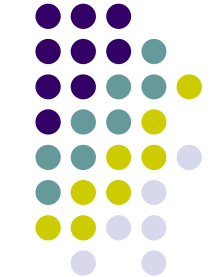
$$\begin{aligned} H_{\mathcal{D}}(F[Q]) &= L(\binom{1}{0}) \cap \left(L(\binom{1}{1}) - 1\right) \cap \left(L(\binom{0}{1}) - 2\right) \\ &= \{4,8\} \cap \{0,2,4,6,8\} \cap \{0,4,8\} \\ &= \{4,8\}. \end{aligned}$$

Indexing and Retrieval



- Large memory requirement for indexing
  - Solution:
    - Break feature set, Fu, Fl, Fm
    - Query individual indexes
    - Post process with additional merging/intersection
- Exact match at feature and segment level while still allowing lot of variability at frame level

Results -Indexing



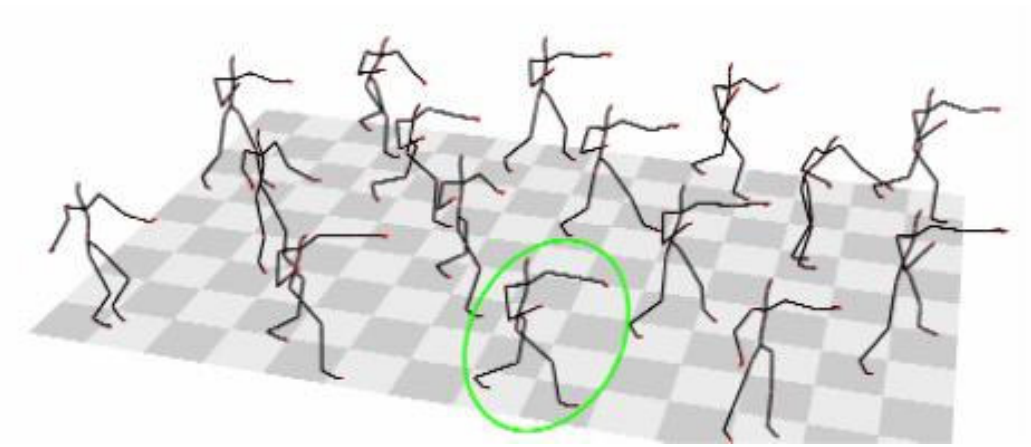
- O(n) in number of frames
- Index size proportional to # segments
- #segments = 3 to 5% of # framesc

Index	f	2^f	#(lists)	#(frames)	#(segs)	MB	$\frac{\text{bytes}}{\text{seg}}$	t_r	t_f	t_i	$\sum t$
I_ℓ^{60}	11	2048	409	425,294	21,108	0.72	35.8	26	10	6	42
I_ℓ^{180}	11	2048	550	1,288,846	41,587	1.41	35.5	71	26	13	110
I_u^{60}	12	4096	642	425,294	53,036	1.71	33.8	26	13	10	49
I_u^{180}	12	4096	877	1,288,846	135,742	4.33	33.4	71	33	25	129
I_m^{60}	8	256	55	425,294	19,067	0.60	33.0	26	20	3	49
I_m^{180}	8	256	75	1,288,846	55,526	1.80	34.0	71	54	12	137

Results -Retrieval

- 10,000 random trials

Type, #(segs)	1–9 hits			10–99 hits			≥ 100 hits		
	μ_h	σ_h	μ_t (ms)	μ_h	σ_h	μ_t (ms)	μ_h	σ_h	μ_t (ms)
exact, $\|Q\| = 5$	3.0	2.4	16	44	28	20	649	567	144
exact, $\|Q\| = 10$	1.7	1.6	17	34	22	26	239	147	71
exact, $\|Q\| = 20$	1.1	0.6	19	32	26	36	130	5	52



Results