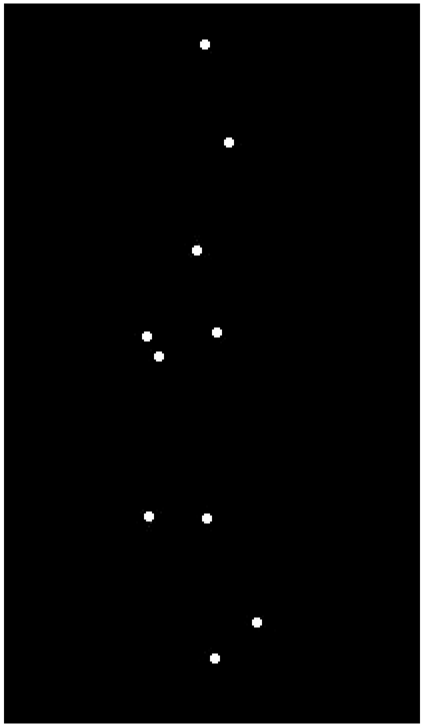
Learning the space of time warping functions for Activity Recognition

Function-Space of an Activity

Ashok Veeraraghavan
Rama Chellappa
Amit K. Roy-Chowdhury

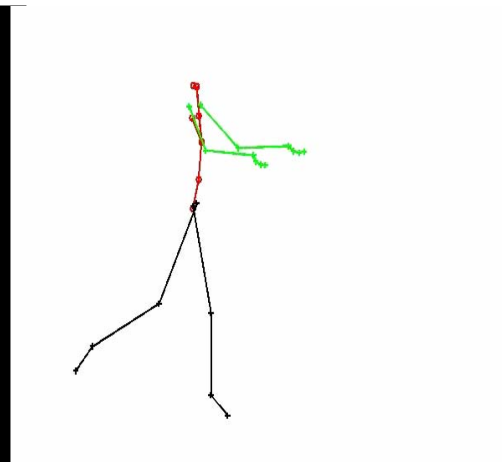
Human Motion Analysis



[Johansson 1973]

3D models of humans

with joint angles



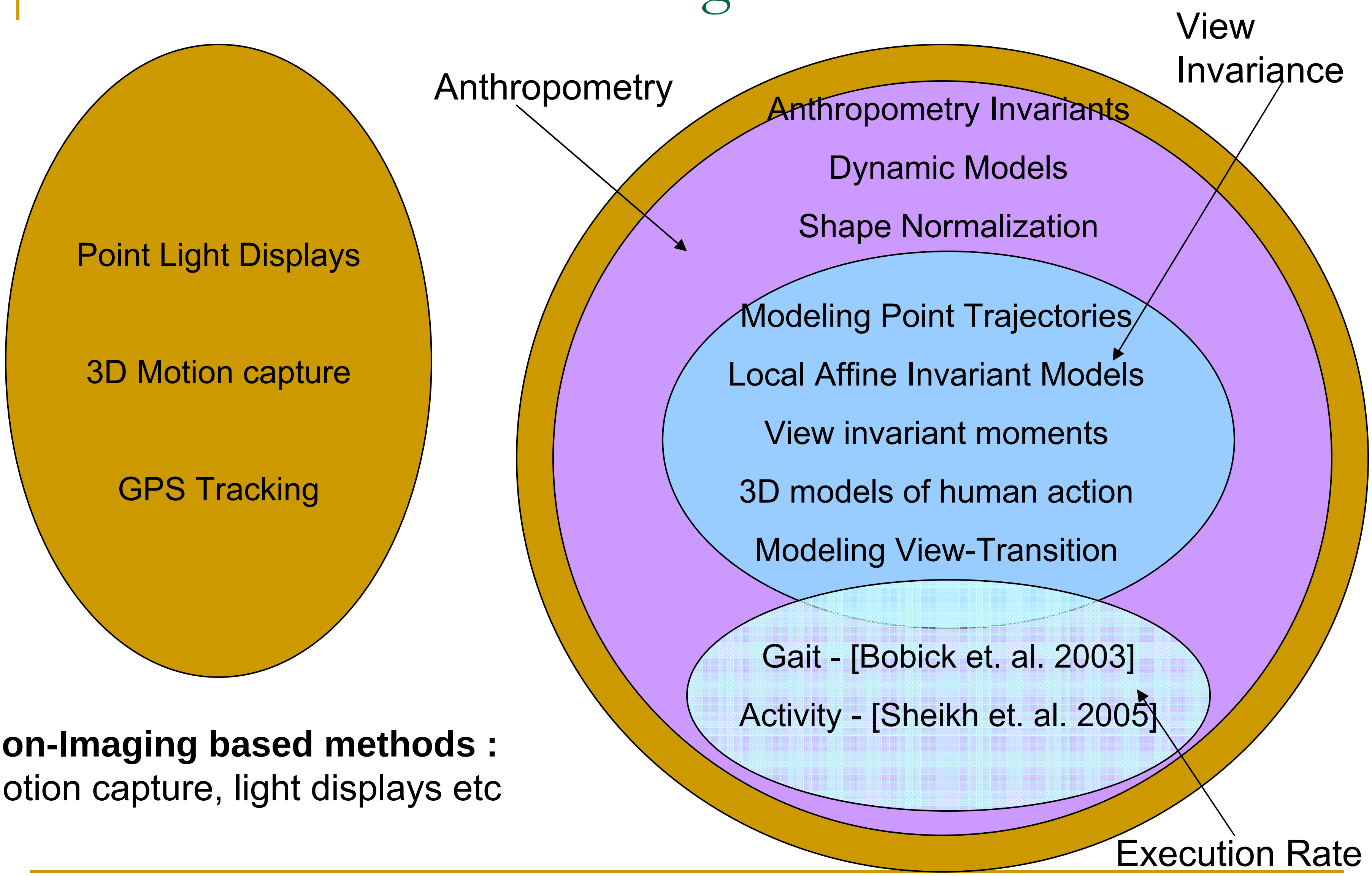
Motion capture

[mocap.cs.cmu.edu]

Shape Sequences

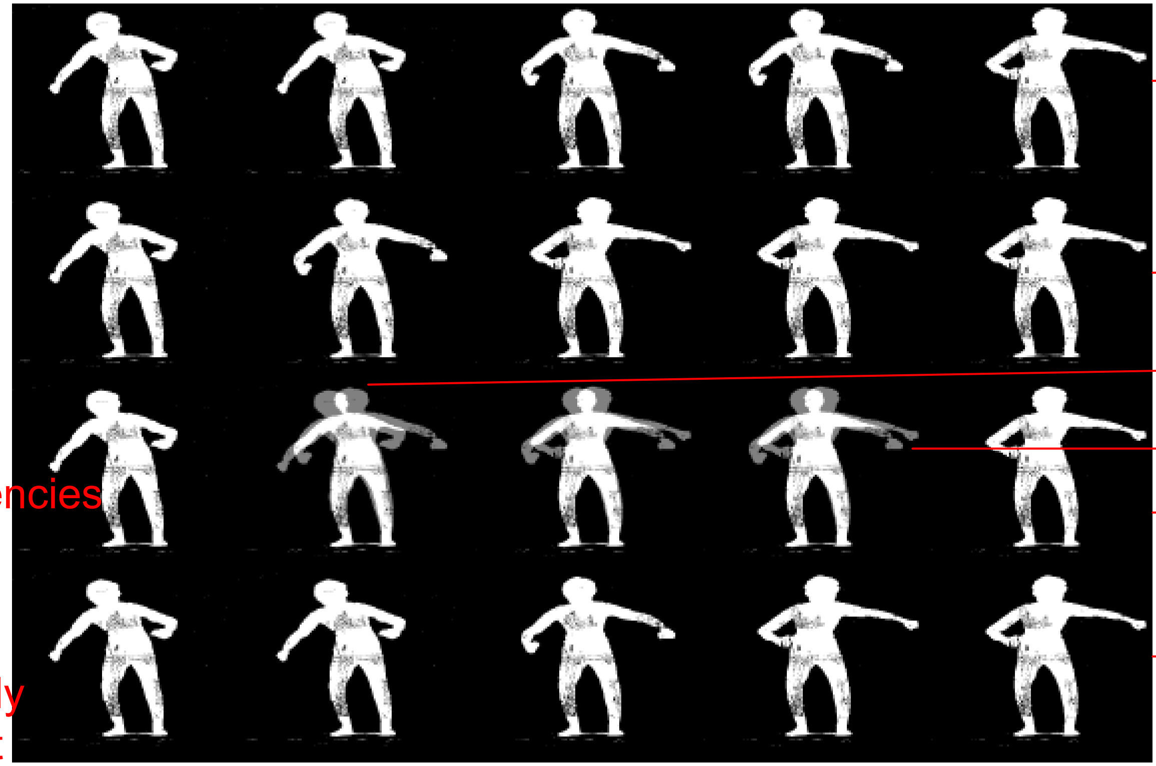
[Veeraraghavan 2004]

Human Action Recognition – Prior Work



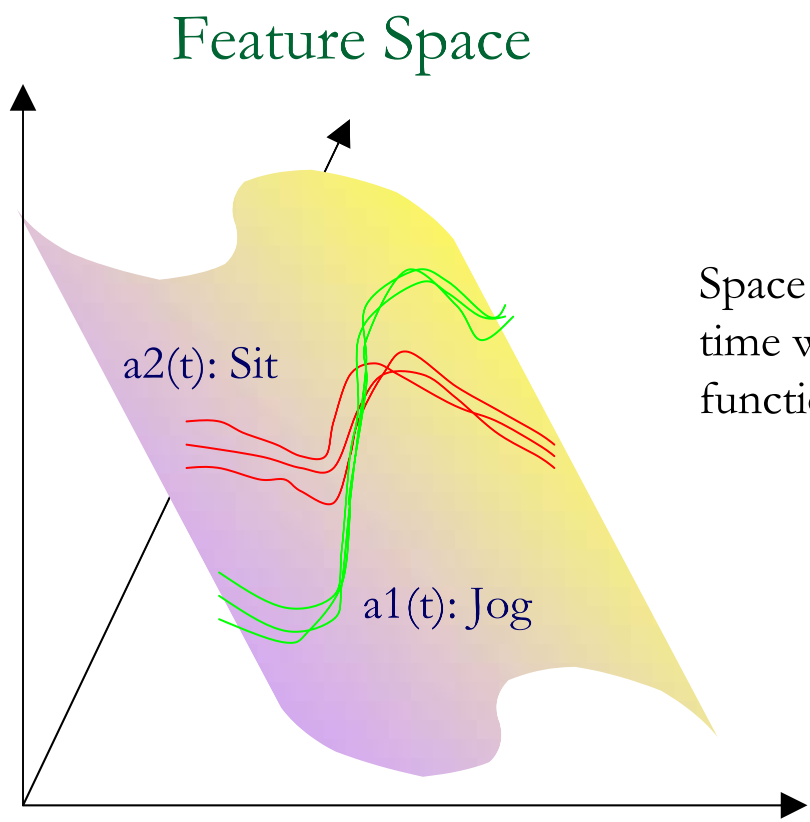
Non-Imaging based methods :
Motion capture, light displays etc

Why is learning Execution rate important?

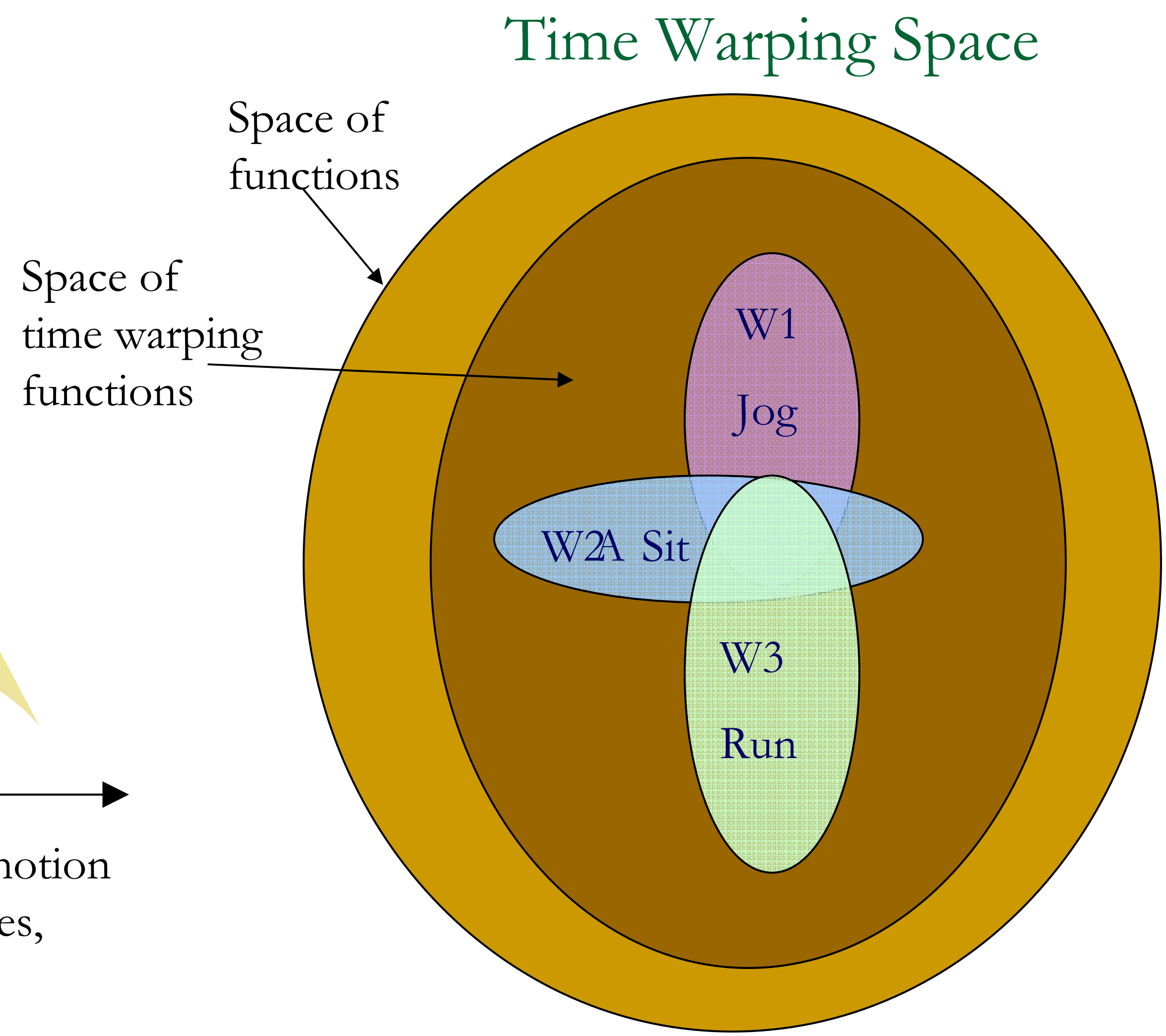


- Structural Inconsistencies. (2 heads, 4 arms etc.)
- Wrong match in recognition experiments.
- Algorithms attempt to explain temporal variation by modeling feature variation.

Model for an Activity



Possible Features: Silhouettes, 3D motion capture, shape of outline, body angles, location of joints in 3D/2D etc



Modeling time-warping for Activities

Model For an Activity

- Nominal activity trajectory : a(t);   t ε (0,1)
- W : Space of time warping functions.
- Realizations of the activity

$$r(t) = a(f(t)) \text{ where } f \in W$$

- Also need to know how to sample candidate functions "f" from W.

Properties of functions in A

1. All realizations of the activity starts at time t=0 and ends at time t=1, i.e., f(0)=0 and f(1)=1.
2. The order of action units for each activity remains unaltered for all realizations i.e.,

$$f'(t) > 0 \; \forall \, t \in (0,1)$$

3. We note that A is a convex set., i.e., if f_1 and f_2 are in A, then for α є (0,1)

$$f = \alpha f_1 + (1-\alpha) f_2$$

f is also in A.

University of California, Riverside

Activity Specific time-warping space W

- W is a subset of A the space of time warping functions.
- $f(t) = t$ is a candidate function in W. This represents no time warping.
- It is reasonable to assume that W is pointwise convex, i.e., for all $f1, f2 \in W$ and $\alpha \in (0,1)$, $f = \alpha f1 + (1-\alpha) f2$ is also in W.
  - Since the derivative is a linear operator, this means that if the rate of execution of some action can be speeded up by factors $\alpha 1$ and $\alpha 2$ then it can also be speeded up by any factor β in between $\alpha 1$ and $\alpha 2$. This is not just reasonable but in fact desirable.

Activity Specific time-warping space W

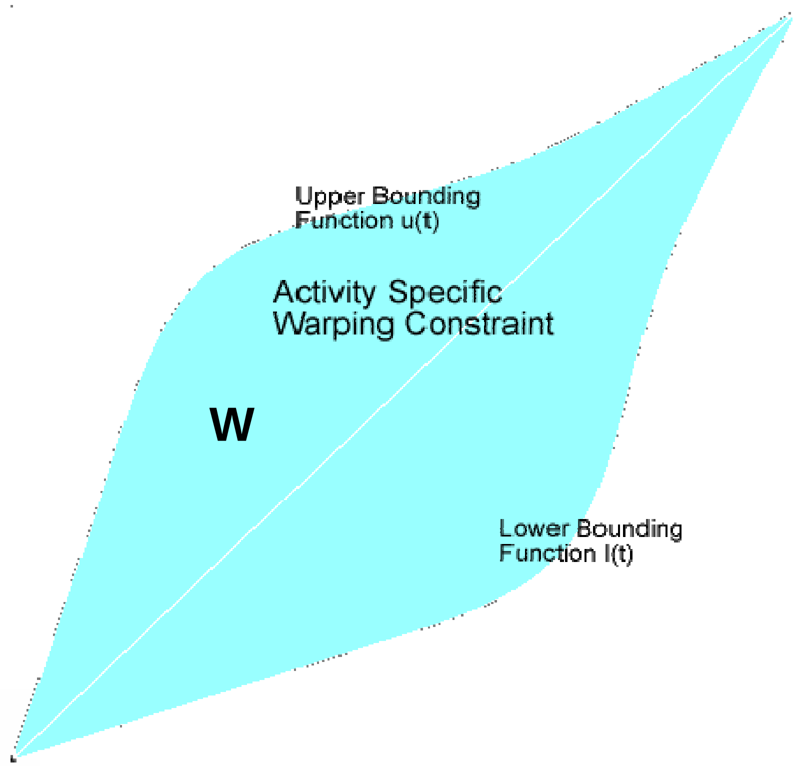
- These properties imply that W can be represented by a warping constraint window, given by two functions u(t) and l(t) where u(t) is the upper bounding function and l(t) is the lower bounding function.

- Time warping functions f in the activity specific warping space W are such that

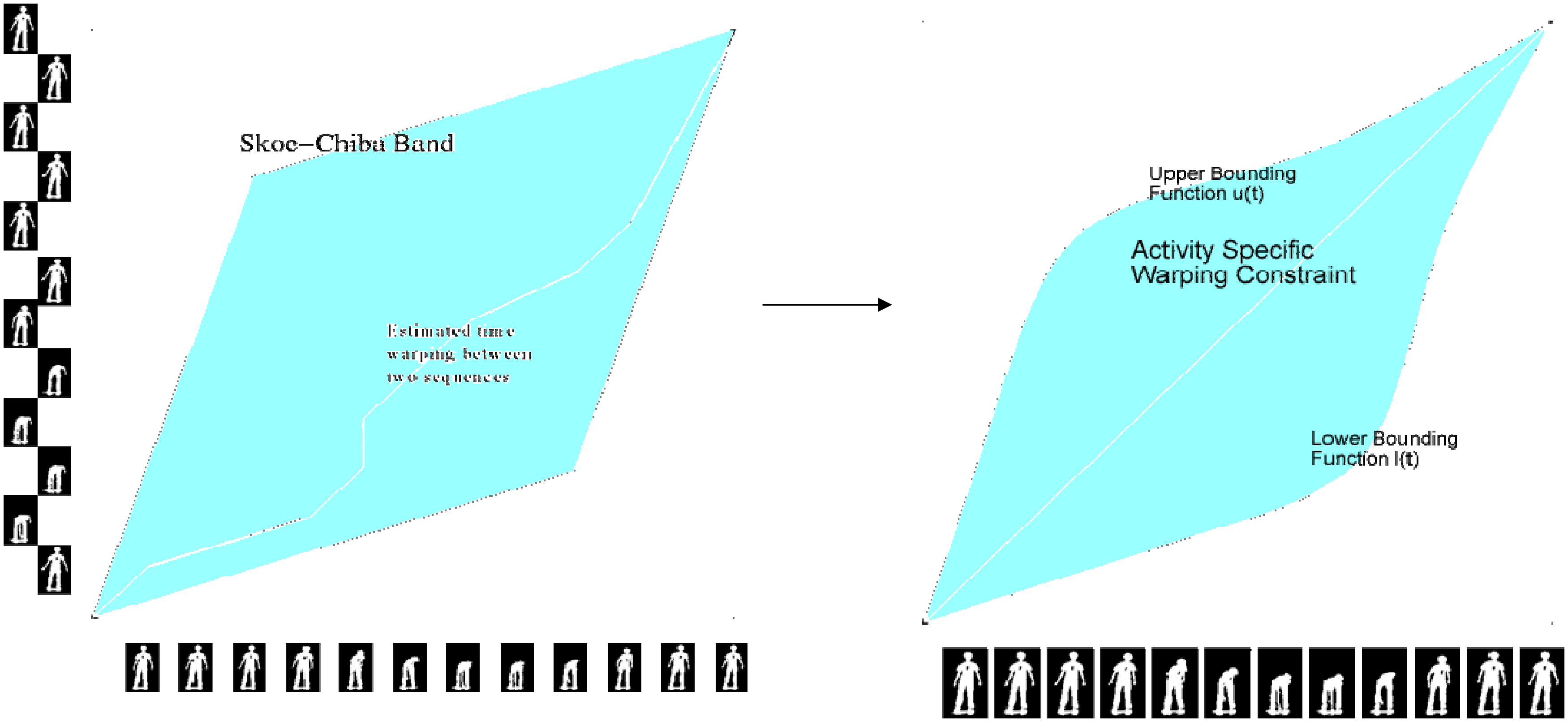
$$u(t) \geq t \geq l(t) \quad \forall t \in (0,1)$$

$$u \geq f \geq l \quad \forall f(t) \in W$$

$$\text{where} \quad f \geq g \Longrightarrow f(t) \geq g(t) \quad \forall t \in (0,1)$$



From DTW to Activity Specific Warping Constraints



Symmetric Representation of Activity Model

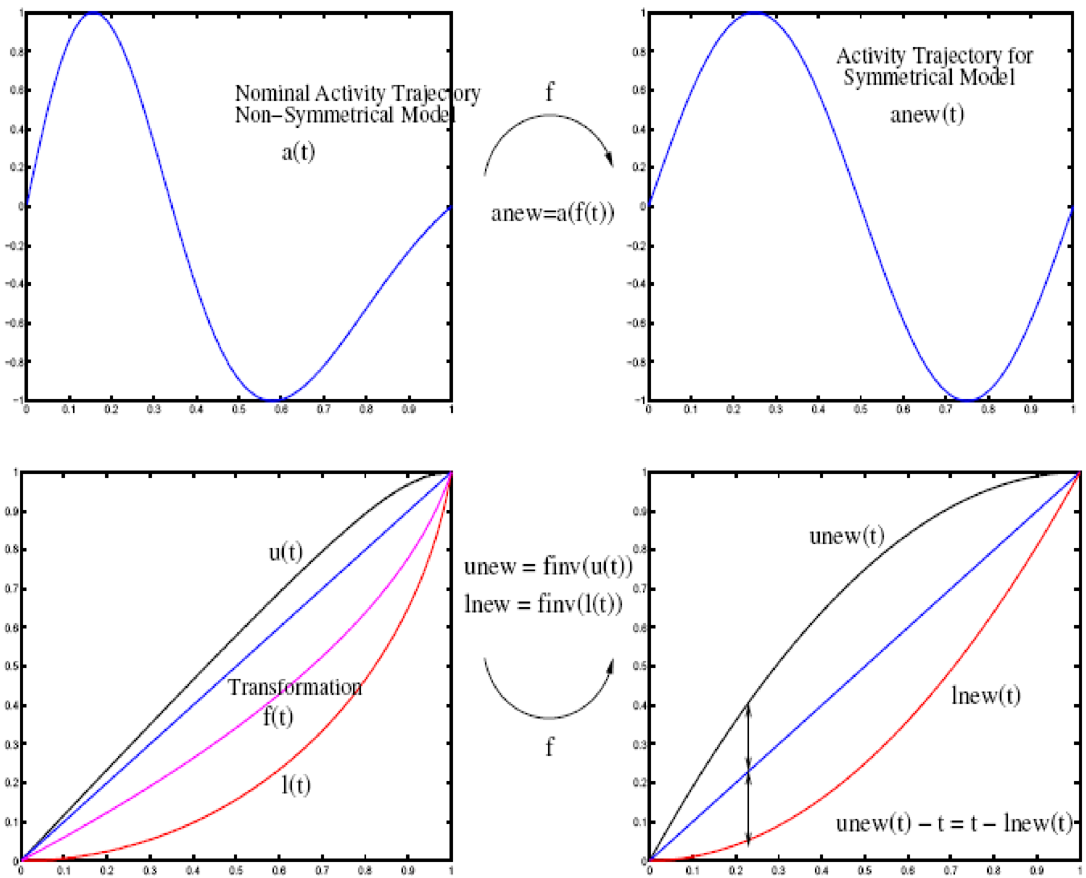


Figure 3. Non symmetrical model $\{a(t), W_{ul}\}$ and the corresponding symmetrical version $\{a_{new(t)}, W_{unew,lnew}\}$. The warping transformation between the two models is given by f.

- Model parameters {a, W} non-unique.
- There exists an equivalence class of model parameters.
- Each equivalence class contains one member whose model parameters are symmetric.
- Learning and inference made only on symmetric representation of model in order to ensure uniqueness.

Learning Symmetric Representation of Activity Model

- Learn the nominal activity trajectory a(t)
- Learn the functional space of time-warps W.
- Learning algorithm can be based on any chosen time alignment procedure.
- We have used the Dynamic Time Warping (DTW) for time alignment.
- EM based learning algorithm
- Expectation

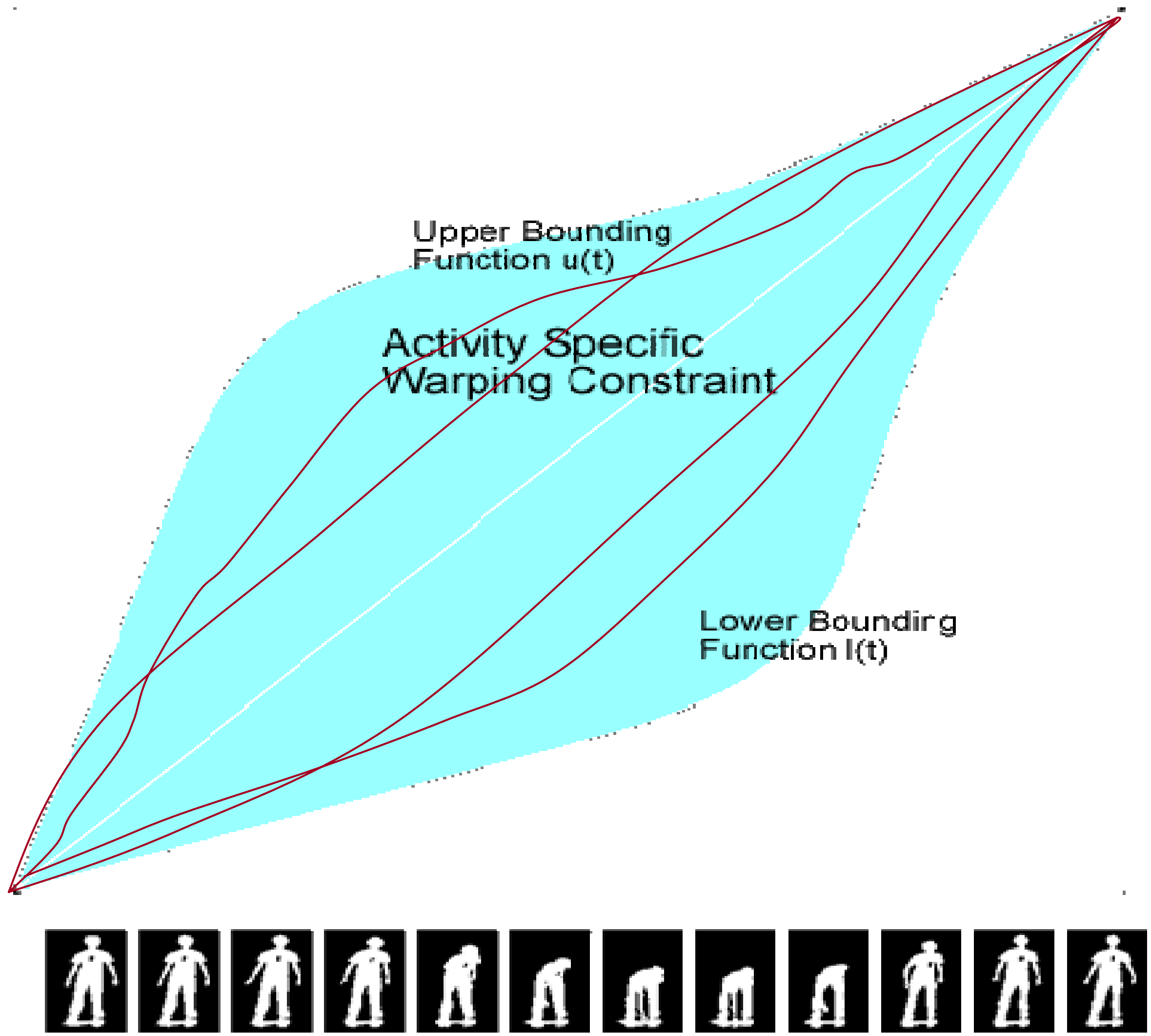
$$\hat{b}(t) = E(b_i(g_i^{-1}(t))) = \frac{1}{N}\sum_{i=1}^{i=N} b_i(g_i^{-1}(t))$$

- Maximization

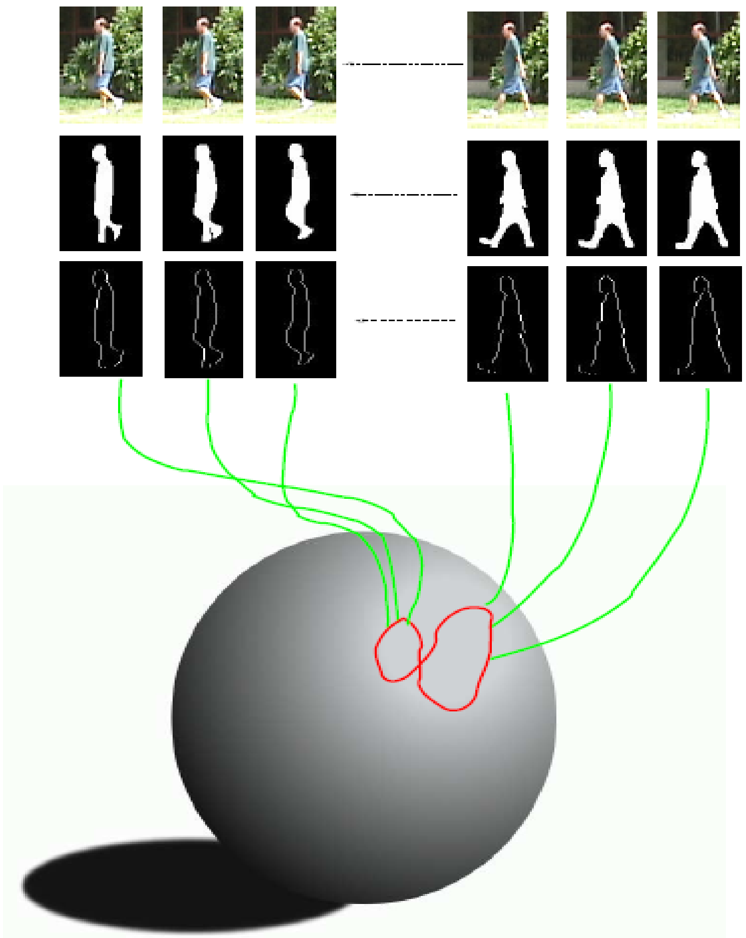
$$\hat{u}(t) = \max_{i=1,2,\ldots N} g_i(t) \ \forall t \in (0,1)$$

$$\hat{l}(t) = \min_{i=1,2,\ldots N} g_i(t) \ \forall t \in (0,1)$$

Learning the Model



Pre-processing and Feature Extraction



[Veeraraghavan et. al.] CVPR 2004, PAMI 2005

1. Background Subtraction
2. Connected Component analysis
3. Extract Shape Feature. (Kendall's statistical shape)
4. Shape Feature lives on a spherical shape-space.
5. Use appropriate distance measures like Procrustes distance for local distance computations.

Results on USF Data

- 70 people, upto 10 sequences per person
- Variabilities: shoe type, surface, view point.

Table 1. Comparison of Identification rates on the USF dataset

Pr-obe	Base-line	DTW Shape	HMM Shape	HMM Image	DTW R-R	Our method
Avg.	42	42	41	50	42	59
A	79	81	80	96	52	70
B	66	74	72	86	52	68
C	56	52	56	74	72	81
D	29	29	22	32	33	40
E	24	20	20	28	26	64
F	30	19	20	17	26	37
G	10	19	19	21	36	53

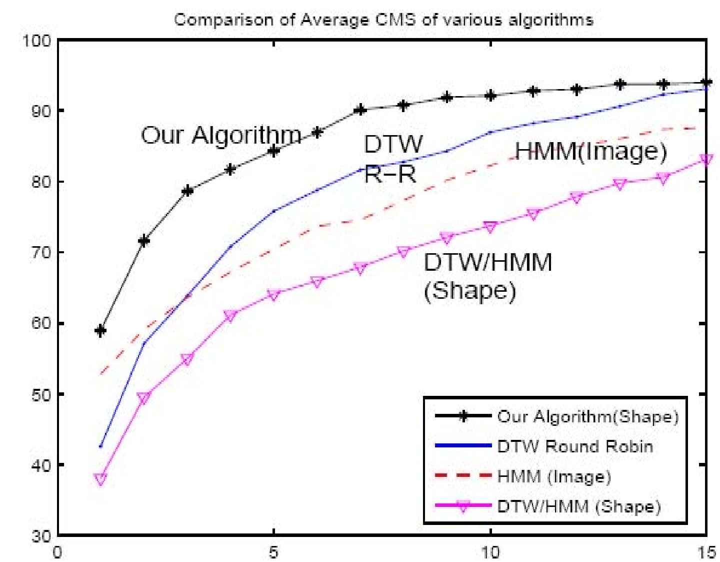
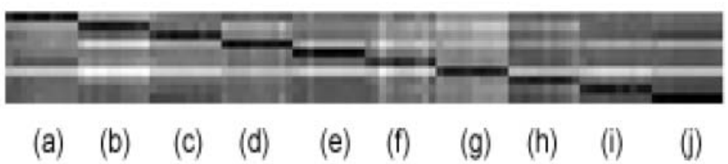


Figure: CMS curve for various algorithms on the USF database.

Baseline : [Sarkar 2005]
DTW Shape, HMM Shape : [Veeraraghavan 2004]
HMM Image : [Kale 2004]

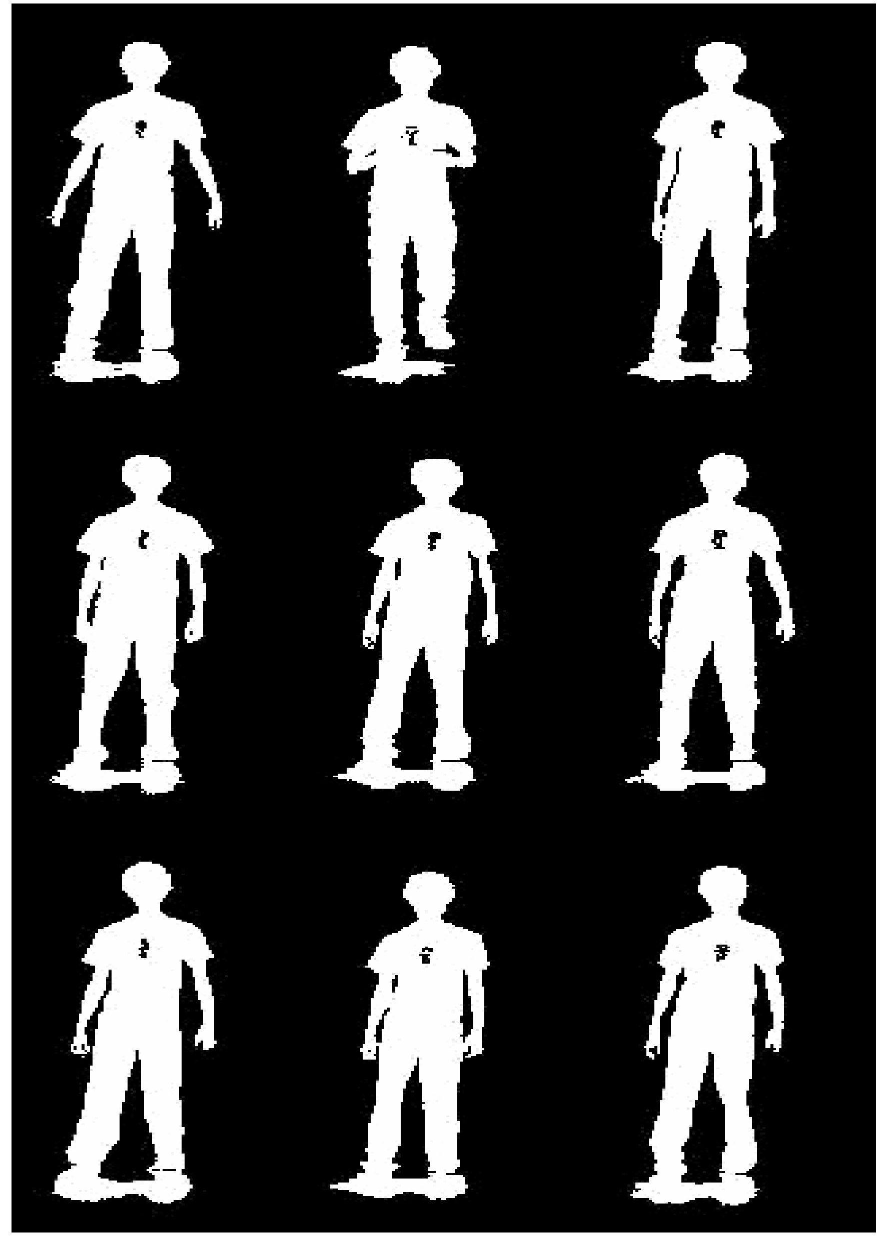
Activity Recognition on UMD Database

1. Database of 10 Activities and 10 Sequences per activity
2. 100 % Recognition and increase in discrimination in the similarity matrix compared to traditional DTW.



(a) Pick up Object
(b) Jog in place
(c) Push
(d) Squat
(e) Wave
(f) Kick
(g) Bend to the side
(h) Throw
(i) Turn Around
(j) Talk on Cellphone

Figure 4. 10 X 100 Similarity matrix of 100 sequences and 10 different activities.



Organize a large Database Hierarchically (Dendrogram)

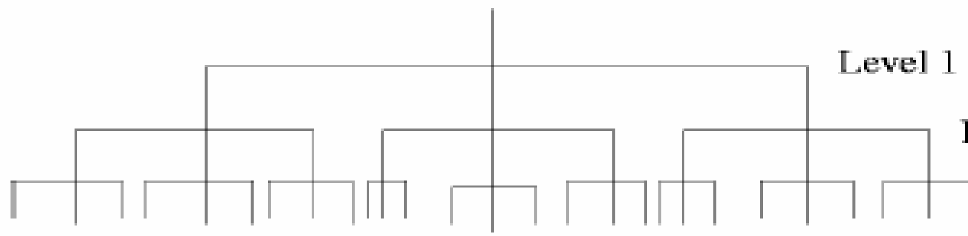


Figure 6. Dendrogram for organizing an activity database

1. USF Database. Total 1870 Sequences of 122 individuals.
2. No: of Leaves at every node = 3.
3. No: of Levels of Dendrogram = 3

$$\eta = 100 * \frac{\text{Identification rate after organization}}{\text{Identification rate before organization}}$$

Table 2. Efficiency of Organization on the USF dataset

Probe	A	B	C	D	E	F	G	Avg
η	76	81	84	100	82	100	95	89

Conclusions

- Modeling, Learning and accounting for time warping is important for activity recognition.
- We proposed a convex activity specific function space for time warping functions and derived learning, recognition and clustering algorithms using this model.
- Appropriate feature selection will enable view and anthropometry invariance.

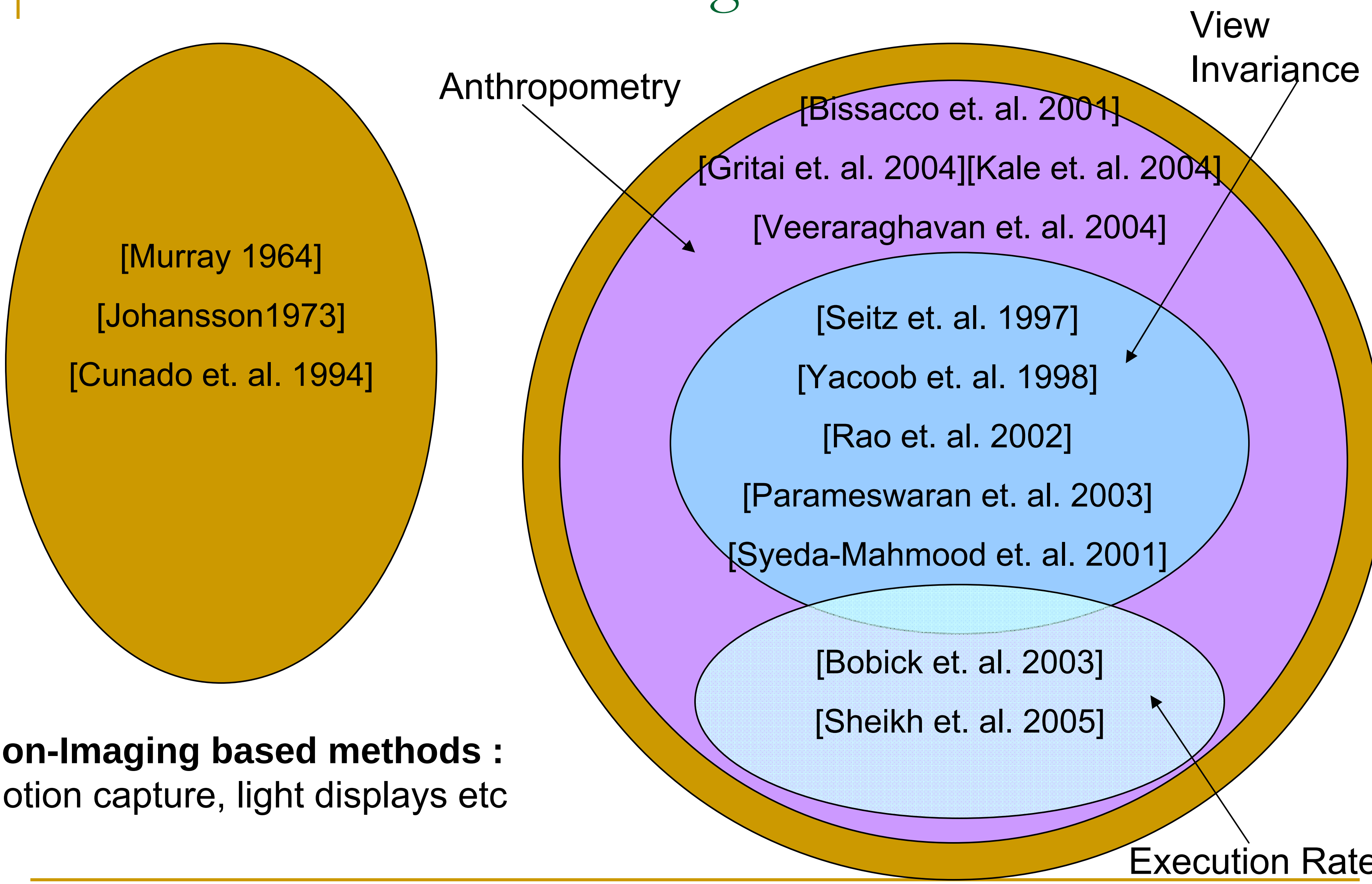
Thank You!

- Contact : vashok@umd.edu

Thank You!

- Contact : vashok@umd.edu

Human Action Recognition – Prior Work



Non-Imaging based methods :
Motion capture, light displays etc