The KidsRoom:

A Perceptually-Based Interactive and Immersive Story Environment

An MIT MediaLab Experiment

A. Bobick, S. Intille, J. Davis, F. Baird,
C. Pinhanez, L. Campbell, Y. Ivanov,
A. Schutte, and A. Wilson

What is a Perceptually-Based Interactive and Immersive Story Environment?

An interactive physical story environment using a computer system which is fully aware of the current state of the room at any given moment in time

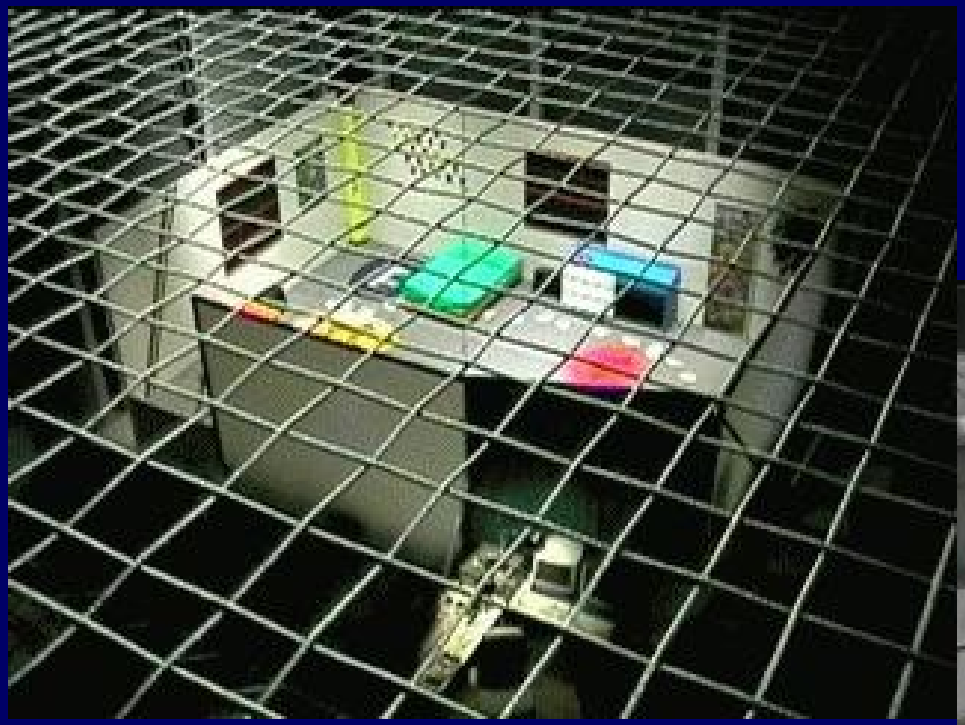
Goals

- Building an “intelligent, aware computer”
- Allowing action to take place in physical space
- Supporting multiple people
- Building a fun environment for children

Goals

- Using vision-based remote sensing
- Constructing an environment that demonstrates various computer vision technologies
- Using context to increase reliability

The Playspace



Wire-grid ceiling, 27 feet high

24 by 18 feet “bedroom” with real furniture
2 projectors

Six computers

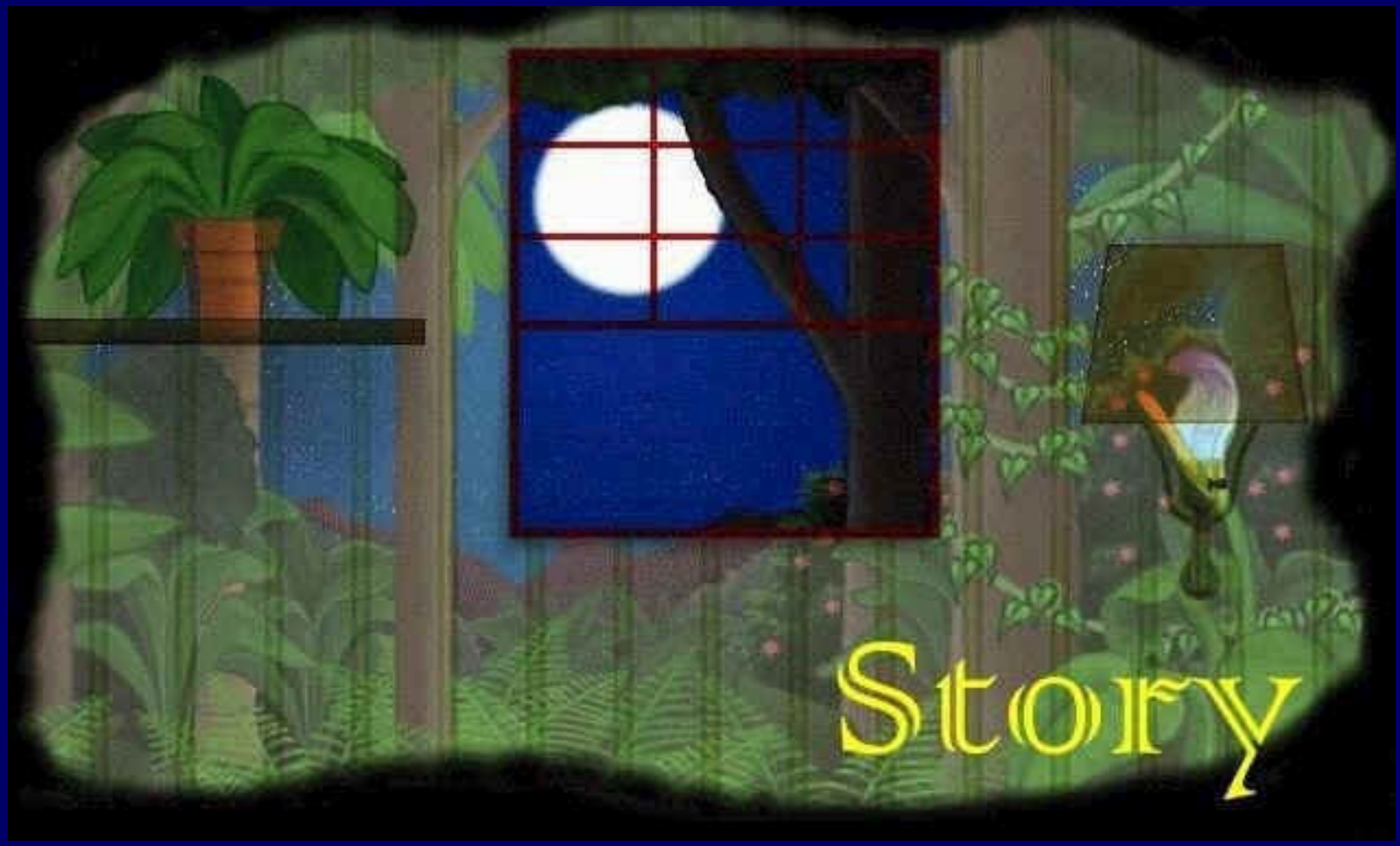
Cameras

Camera 1:
Top View
Used for tracking people in all worlds and for detecting rowing in the river world.

Camera 3:
Red Rug
Used for action recognition during the monster dance.

Camera 2:
Green Rug
Used for action recognition during the monster dance.

Camera 4:
Spectator View



“Magical World”
Linear narrative (no branching story lines)
Reactive interaction
Strong motivation for group behavior

The Four Worlds

The Bedroom World

The Forest World

The River World

The Monster World

The Bedroom World

- Children enter one by one
- Has appearance of a child's bedroom:
  Bed, 2 rugs, desk
- Scavenger hunt for magic word: children send from one piece of furniture to the next

randomness

The Forest World

- Transition from bedroom world when magic word found: lights change
- “Follow the path”
- “Monsters are near, stay in group”
- “Hide behind the bed”

The River World

"The magic bed is now a boat"

- "Passenger overboard!"
- "Row and watch out!"
- "You made it! Push the boat on shore."

The Monster World

- “Yell! Keep the monsters away!”
- “Let’s Dance!”
- One child per rug (non-occluded view of child)
- Four dances
- Interaction with animated monste monitor
- Imitation

Object Tracking

- Tracks up to 4 people and bed using overhead camera
  - -Collects positional information
  - -Scavenger hunt
- Users modeled as 2D background-difference blobs

How Does the Intelligent Environment Interpret Motion?

The pixel-by-pixel difference between consecutive frames is aggregated as the "rowing energy" or "dance energy."

Image differencing: Subtract previous frame from current frame, pixel by pixel:

$$Diff(x,y,t) = I(x,y,t) - I(x,y,t-1)$$

Motion Energy Image

1. Image differencing: Subtract previous frame from current frame, pixel by pixel:

   $$Diff(x,y,t) = I(x,y,t) - I(x,y,t-1)$$

2. Aggregate T difference images into a single binary "motion energy image" $B_{energy}(x,y,t)$.
3. Use T> 20, so that you can really see the difference.

Motion Energy Definition

The union operator U creates a binary image: 1 for any pixel for which Diff>0 in any of the T frames, 0 otherwise:

$$B_{energy}(x,y,t) = \bigcup_{i=0}^{T} Diff\,(x,y,t-i)$$

The Y Dance

Idea: Measure shape of

“motion blob”

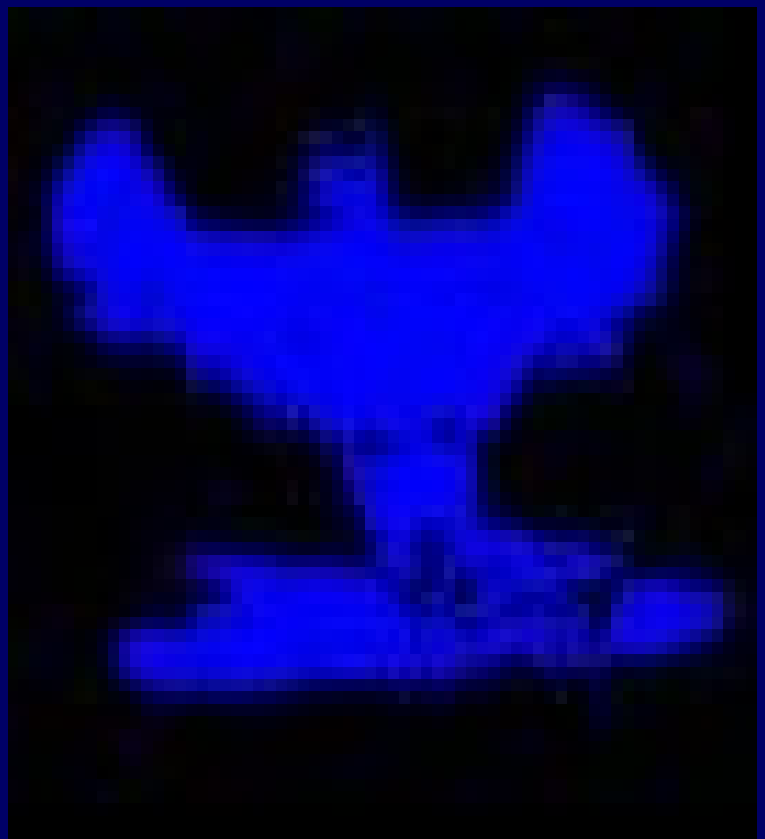
=

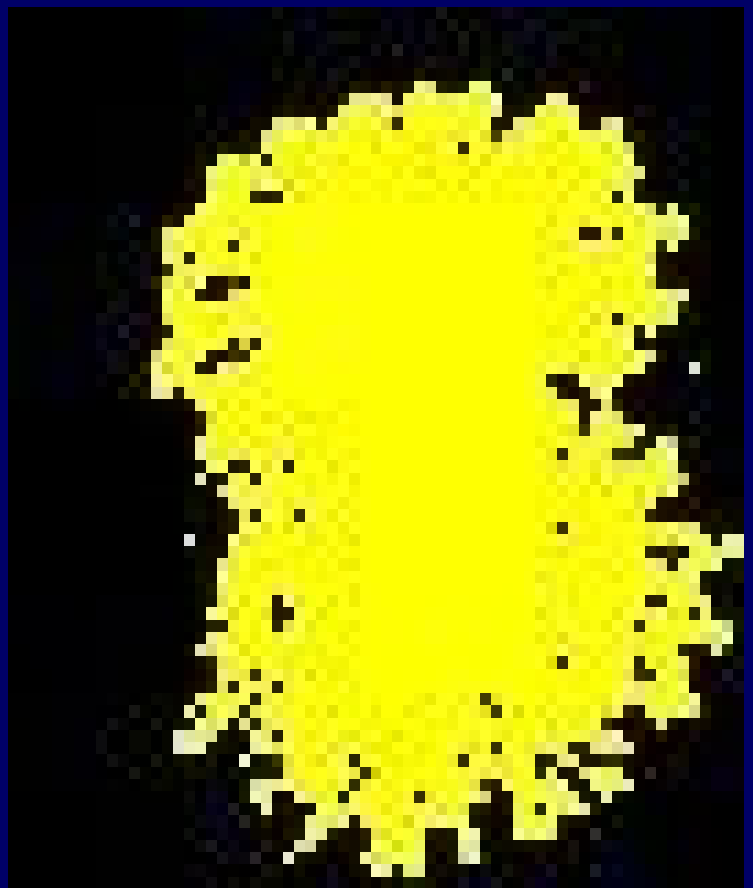
“background difference blob”

= “motion energy blob”

The Wing Flap and Spin Dances

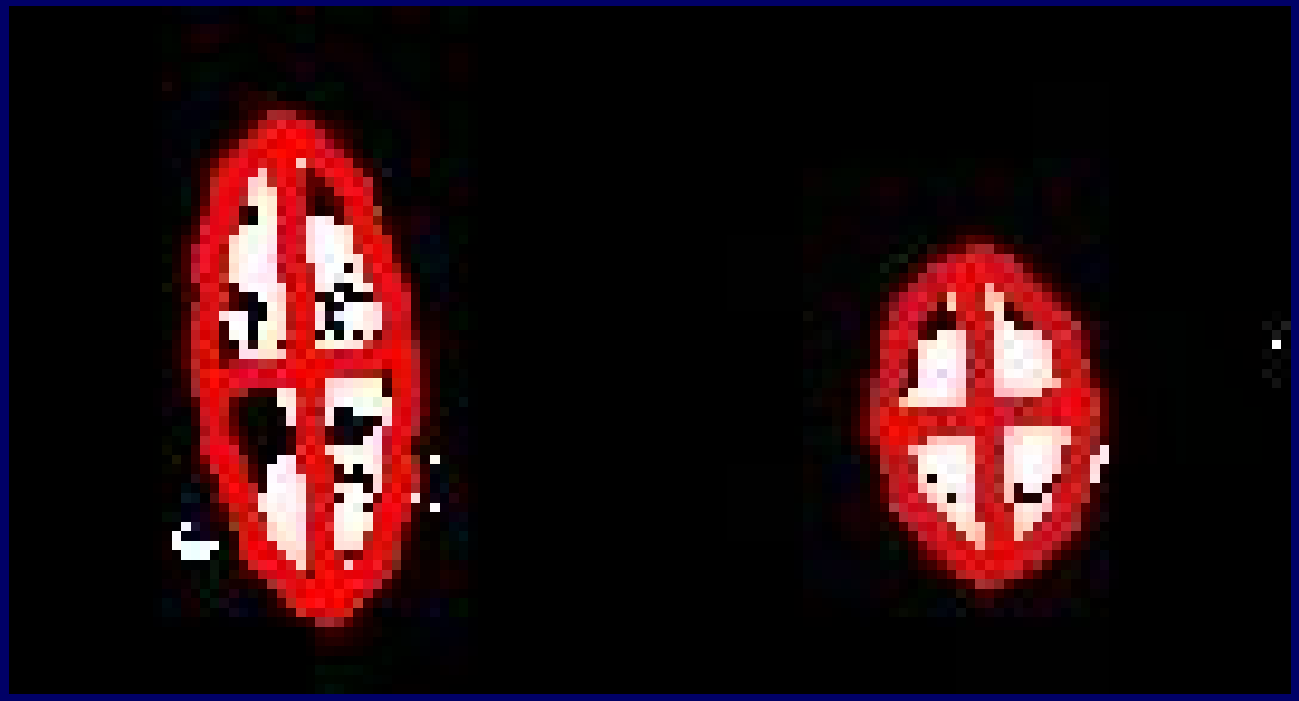
Match Temporal Templates:





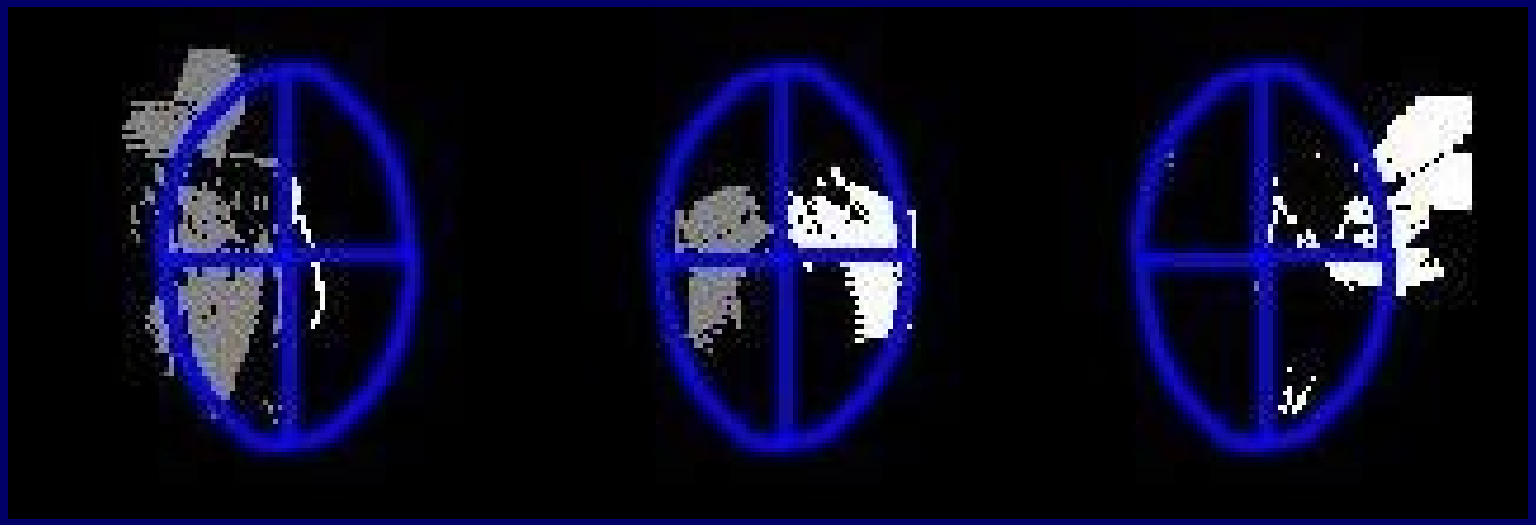
The Crouching Dance

Measure size of background difference blob:



Rowing - Motion Energy

- Ellipse = position and orientation of bed
- The pixel-by-pixel difference between consecutive frames is summed up as the “rowing energy.“
- Large difference between frames = “high energy”



Lessons and Observations

- Importance of context for action recognition
- Importance of a story to make participants cooperate
- Importance of having the algorithms fail gracefully to maintain realism