

# Representing Moving Images with Layers

CSE 455

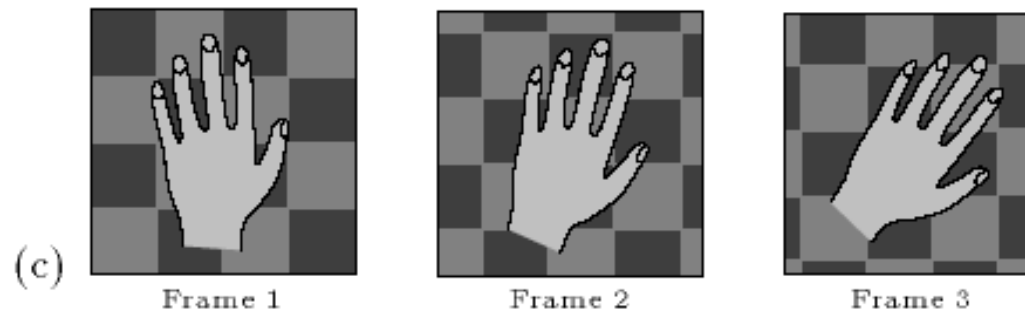
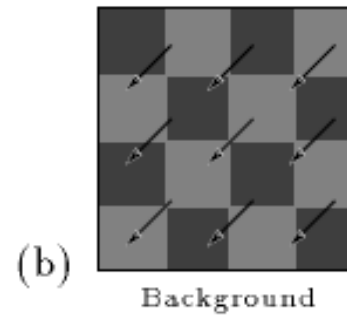
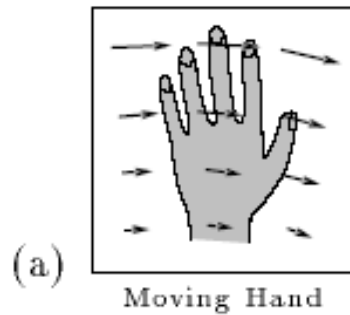
Linda Shapiro

from J. Y. Wang and E. H. Adelson  
MIT Media Lab

# Goal

- Represent moving images with sets of overlapping layers
- Layers are ordered in depth and occlude each other
- Velocity maps indicate how the layers are to be warped over time

# Simple Domain: Gesture Recognition



# More Complex: What are the layers?



# Even More Complex: How many layers are there?

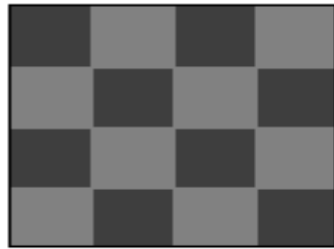


# Definition of Layer Maps

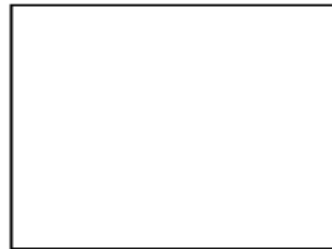
- Each layer contains three maps
  1. **intensity map** (or texture map)
  2. **alpha map** (opacity at each point)
  3. **velocity map** (warping over time)
- Layers are ordered by depth
- This can be for vision or graphics or both

# Layers for the Hand Gestures

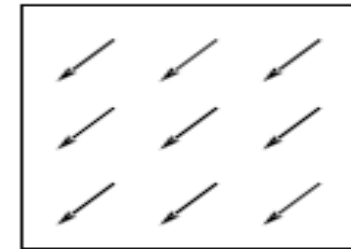
Background  
Layer



Intensity map

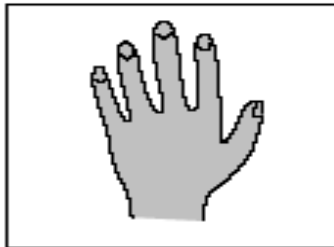


Alpha map



Velocity map

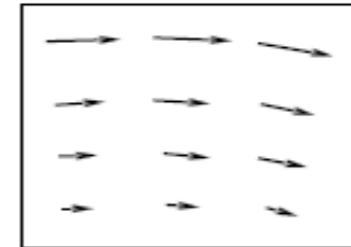
Hand Layer



Intensity map

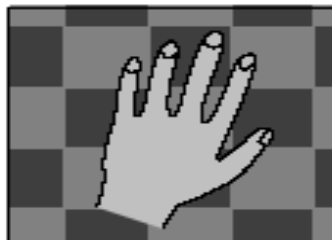


Alpha map



Velocity map

Re-synthesized  
Sequence



# Optical Flow Doesn't Work

- Optical flow techniques typically model the world as a 2-D rubber sheet that is distorted over time.
- When one object moves in front of another, the rubber sheet model fails.
- Image information appears and disappears; optical flow can't represent this.
- Motion estimates near boundaries are bad.



# Block Matching Can't Do It

- Block motion only handles translation well.
- Like optical flow, block matching doesn't deal with occlusion or objects that suddenly appear or disappear off the edge of the frame.

# Layered Representation: Compositing

$$I_1(x, y) = E_0(x, y)(1 - \alpha_1(x, y)) + E_1(x, y)\alpha_1(x, y). \quad (1)$$

- $E_0$  is the background layer.
- $E_1$  is the next layer (and there can be more).
- $\alpha_1$  is the alpha channel of  $E_1$ , with values between 0 and 1 (for graphics).
- The velocity map tells us how to warp the frames over time.
- The intensity map and alpha map are warped together, so they stay registered.

# Analysis: Flower Garden Sequence

Frame 1



(a)

Frame 15



(b)

Frame 30



(c)

Camera is panning to the right.

Figure 6: Frames 0, 15 and 30, of MPEG flower garden sequence are shown in figures (a-c), respectively.



(a)

Frame 1  
warped



(b)

Frame 15

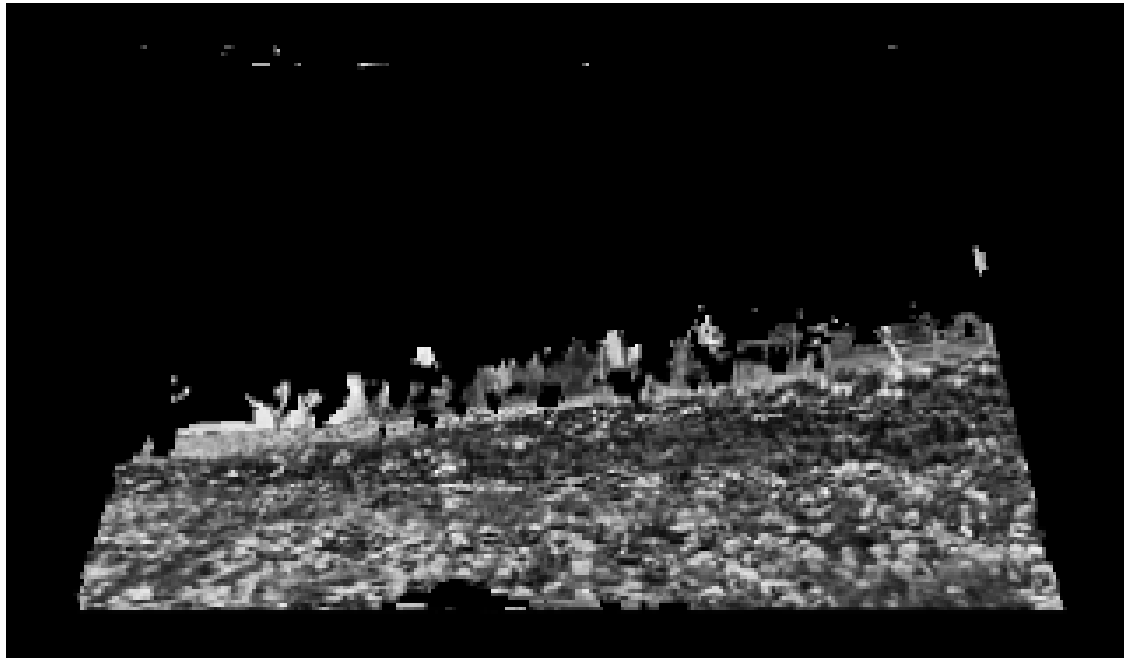


(c)

Frame 30  
warped

What's going on here?

# Accumulation of the Flowerbed Layer



# Motion Analysis

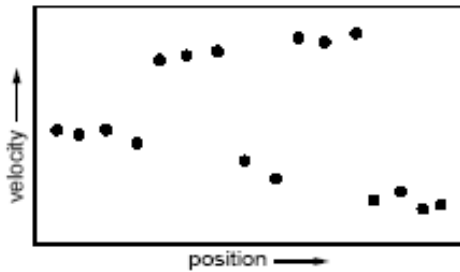
1. Robust motion segmentation using a parametric (affine) model.

$$V_x(x,y) = a_{x0} + a_{xx}x + a_{xy}y$$

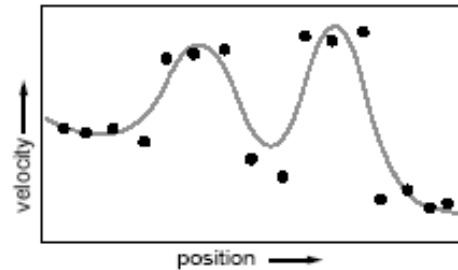
$$V_y(x,y) = a_{y0} + a_{yx}x + a_{yy}y$$

2. Synthesis of the layered representation.

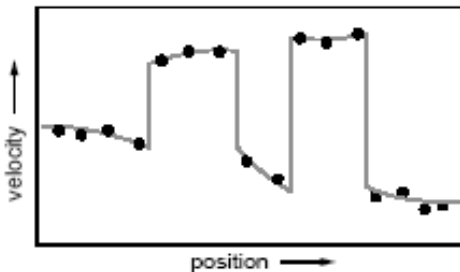
# Motion Analysis Example



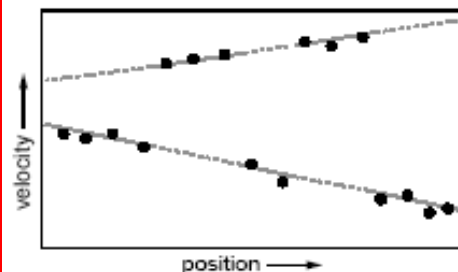
(a) velocity estimates



(b) velocity smoothing



(c) regularization



(d) robust estimation

2 separate layers  
shown as 2 affine  
models (lines);

The gaps show  
the occlusion.

# Motion Estimation Steps

1. Conventional optical flow algorithm and representation (uses multi-scale, coarse-to-fine Lucas-Kanade approach).
2. From the optical flow representation, determine a set of affine motions. Segment into regions with an affine motion within each region.

# Motion Segmentation

1. Use an array of non-overlapping **square regions** to derive an initial set of motion models.
2. Estimate the **affine parameters** within these regions by linear regression, applied separately on each velocity component ( $dx$ ,  $dy$ ).
3. Compute the **reliability** of each hypothesis according to its **residual error**.
4. Use an **adaptive k-means clustering** that merges two clusters when the distance between their centers is smaller than a threshold to produce a set of likely affine models.



# Region Assignment by Hypothesis Testing

- Use the motion models derived from the motion segmentation step to identify the coherent regions.
- Do this by minimizing an error (distortion) function:

$$G(i(x,y)) = \sum_{x,y} (V(x,y) - V_{ai}(x,y))^2$$

where  $i(x,y)$  is the model assigned to pixel  $(x,y)$   
and  $V_{ai}(x,y)$  is the affine motion for that model.

- The error is minimized at each pixel to give the best model for that pixel position.
- Pixels with too high error are not assigned to models.

# Iterative Algorithm

- The initial segmentation step uses an array of square regions.
- At each iteration, the segmentation becomes more accurate, because the parameter estimation is within a single coherent motion region.
- A region splitter separates disjoint regions.
- A filter eliminates small regions.
- At the end, intensity is used to match unassigned pixels to the best adjacent region.

# Layer Synthesis

- The information from a longer sequence must be combined over time, to accumulate each layer.
- The transforms for each layer are used to warp its pixels to align a set of frames.
- The median value of a pixel over the set is used for the layer.
- Occlusion relationships are determined.

# Results

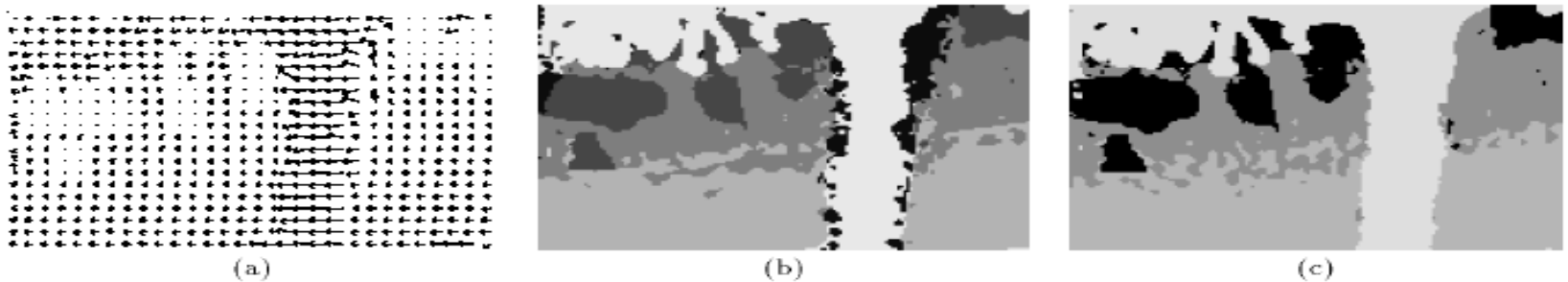


Figure 11: (a) The optic flow from multi-scale gradient method. (b) Segmentation obtained by clustering optic flow into affine motion regions. (c) Segmentation from consistency checking by image warping. Representing moving images with layers.

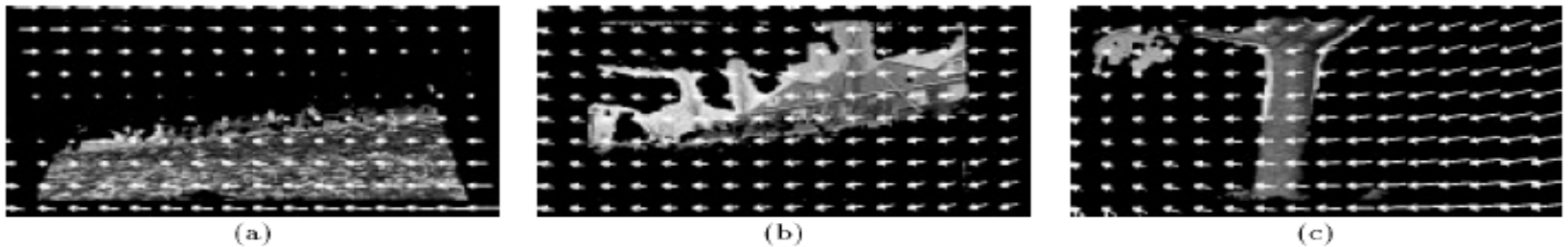


Figure 12: The layers corresponding to the tree, the flower bed, and the house shown in figures (a-c), respectively. The affine flow field for each layer is superimposed.

---

# Results



(a)



(b)



(c)

Figure 13: Frames 0, 15, and 30 as reconstructed from the layered representation shown in figures (a-c), respectively.



(a)



(b)



(c)

Figure 14: The sequence reconstructed without the tree layer shown in figures (a-c), respectively.

# Results

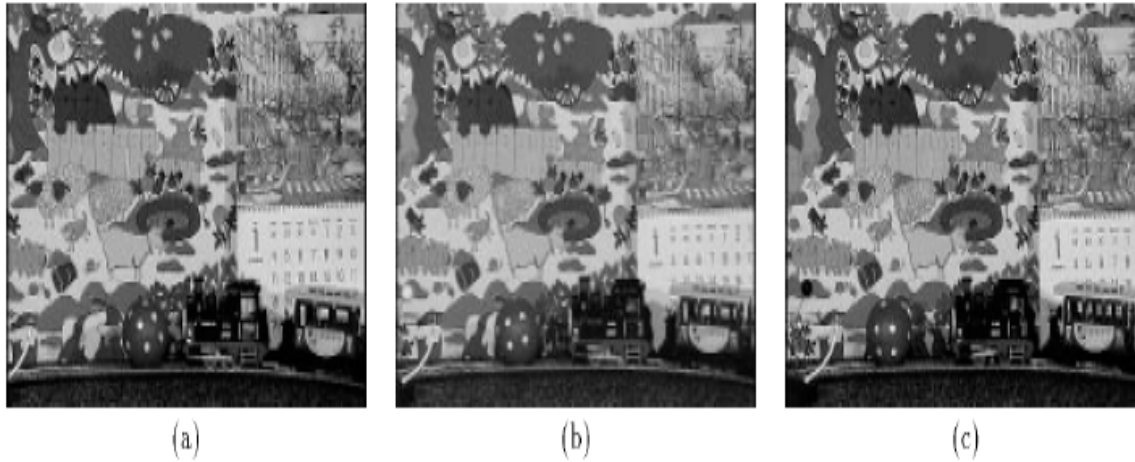


Figure 15: Frames 0, 15 and 30, of MPEG Calendar sequence shown in figures (a-c), respectively.

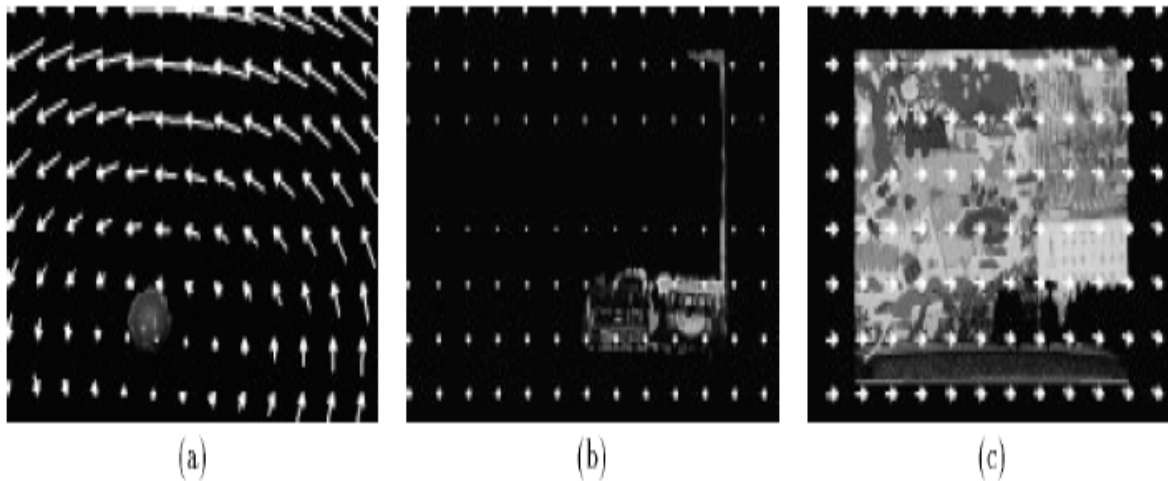


Figure 16: The layers corresponding to the ball, the train, and the background shown in figures (a-c), respectively.