Location and Object Awareness in First Person Vision

Antonino Furnari & Giovanni Maria Farinella

http://www.dmi.unict.it/~furnari/
http://iplab.dmi.unict.it/ftp/
tutorial on
Active Vision and Human Robot Collaboration
Third Person Vision (TPV) vs First Person Vision (FPV)

FIXED CAMERA

✓ Easy to setup
✓ Controlled Field of View
× Doesn’t always see everything
× Not really portable

AGENT-WORN CAMERA

✓ Content is always relevant
✓ Intrinsically mobile
× High variability
× Operational constraints
First Person Vision vs Active Vision

First Person Vision is not strictly active vision (the system cannot manipulate the point of view, the user does it);

➢ FPV allows us to gain a better insight into what the user is doing;
Location and Object Awareness

Location Recognition

The user acquires a 10-seconds video-shot of the personal location

Living Room
Kitchen
Garage
Negative

Next-Active-Object Prediction
Recognizing Personal Locations

Monitoring locations which are important for the user.

The user acquires a 10-seconds video-shot of the personal location

MODEL

new FPV image

user-defined set of locations

Benchmark - Cameras and Representations

• Wearable cameras:
  • Head mounted:
  • Ear mounted:
  • Ear mounted (wide):
  • Chest mounted (wide):

• Representations:
  • Holistic (GIST):
  • Shallow (Bag Of Visual Words - IFV):
  • Deep (CNNs):

Dataset

We acquired a dataset of 8 personal locations arising from possible daily activities.

For each considered device we acquired:

- **Training set**: 10-seconds videos (one for each location);
- **Test set**: five medium-length (5 minutes) videos for each location;

**Classification Models**

**BASELINE**


**PROPOSED**

Entropy-Based (EB) Rejection of Negative Samples

Given $n$ neighboring frames, we assume that they belong to the same class and are conditionally independent with respect to it.

CONTINUOUS PROPERTY OF FPV

Entropy-Based (EB) Rejection of Negative Samples

• Compute posterior probabilities $p(c_i|x_i)$ and combine them:

$$p(c_k|x_1, \ldots, x_n) = \frac{\prod_i p(c_k|x_i)}{\sum_k \prod_i p(c_k|x_i)}$$ (1)

• If the frames are «positive» samples, then the model will agree on their identity and (1) will have low entropy.

• If the frames are «negative» samples, the model will likely give random predictions for different frames and (1) will have high entropy.

threshold on:

$$e(p; x_1, \ldots, x_n) = \sum_k p(c_k|x_1, \ldots, x_k)\log(p(c_k|x_1, \ldots, x_k))$$
### Results (partial)

<table>
<thead>
<tr>
<th>Method</th>
<th>Dev.</th>
<th>Options</th>
<th>Dim.</th>
<th>Proposed</th>
<th>Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_2$</td>
<td>GIST</td>
<td>LX2P</td>
<td>—</td>
<td>512</td>
<td>29.44</td>
</tr>
<tr>
<td>$b_2$</td>
<td>IFV</td>
<td>LX2P</td>
<td>256</td>
<td>40960</td>
<td>17.50</td>
</tr>
<tr>
<td>$c_2$</td>
<td>IFV</td>
<td>LX2P</td>
<td>256 SE</td>
<td>41984</td>
<td>12.56</td>
</tr>
<tr>
<td>$d_2$</td>
<td>IFV</td>
<td>LX2P</td>
<td>512</td>
<td>81920</td>
<td>18.50</td>
</tr>
<tr>
<td>$e_2$</td>
<td>IFV</td>
<td>LX2P</td>
<td>512 SE</td>
<td>83968</td>
<td>18.00</td>
</tr>
<tr>
<td>$f_2$</td>
<td>CNN</td>
<td>LX2P</td>
<td>AlexNet I</td>
<td>4096</td>
<td>70.06</td>
</tr>
<tr>
<td>$g_2$</td>
<td>CNN</td>
<td>LX2P</td>
<td>AlexNet P</td>
<td>4096</td>
<td>64.11</td>
</tr>
<tr>
<td>$h_2$</td>
<td>CNN</td>
<td>LX2P</td>
<td>VGG16 I</td>
<td>4096</td>
<td>67.28</td>
</tr>
<tr>
<td>$i_2$</td>
<td>CNN</td>
<td>LX2P</td>
<td>VGG16 P</td>
<td>4096</td>
<td>63.33</td>
</tr>
<tr>
<td>$j_2$</td>
<td>CNN</td>
<td>LX2P</td>
<td>AlexNet I FT</td>
<td>4096</td>
<td>74.83</td>
</tr>
<tr>
<td>$k_2$</td>
<td>CNN</td>
<td>LX2P</td>
<td>AlexNet P FT</td>
<td>4096</td>
<td>69.94</td>
</tr>
<tr>
<td>$l_2$</td>
<td>CNN</td>
<td>LX2P</td>
<td>VGG16 I FT</td>
<td>4096</td>
<td>68.28</td>
</tr>
<tr>
<td>$m_2$</td>
<td>CNN</td>
<td>LX2P</td>
<td>VGG16 P FT</td>
<td>4096</td>
<td>80.06</td>
</tr>
<tr>
<td>$a_3$</td>
<td>GIST</td>
<td>LX2W</td>
<td>—</td>
<td>512</td>
<td>39.83</td>
</tr>
<tr>
<td>$b_3$</td>
<td>IFV</td>
<td>LX2W</td>
<td>256</td>
<td>40960</td>
<td>37.50</td>
</tr>
<tr>
<td>$c_3$</td>
<td>IFV</td>
<td>LX2W</td>
<td>256 SE</td>
<td>41984</td>
<td>42.83</td>
</tr>
<tr>
<td>$d_3$</td>
<td>IFV</td>
<td>LX2W</td>
<td>512</td>
<td>81920</td>
<td>39.50</td>
</tr>
<tr>
<td>$e_3$</td>
<td>IFV</td>
<td>LX2W</td>
<td>512 SE</td>
<td>83968</td>
<td>37.06</td>
</tr>
<tr>
<td>$f_3$</td>
<td>CNN</td>
<td>LX2W</td>
<td>AlexNet I</td>
<td>4096</td>
<td>75.22</td>
</tr>
<tr>
<td>$g_3$</td>
<td>CNN</td>
<td>LX2W</td>
<td>AlexNet P</td>
<td>4096</td>
<td>73.89</td>
</tr>
<tr>
<td>$h_3$</td>
<td>CNN</td>
<td>LX2W</td>
<td>VGG16 I</td>
<td>4096</td>
<td>70.89</td>
</tr>
<tr>
<td>$i_3$</td>
<td>CNN</td>
<td>LX2W</td>
<td>VGG16 P</td>
<td>4096</td>
<td>81.67</td>
</tr>
<tr>
<td>$j_3$</td>
<td>CNN</td>
<td>LX2W</td>
<td>AlexNet I FT</td>
<td>4096</td>
<td>73.89</td>
</tr>
<tr>
<td>$k_3$</td>
<td>CNN</td>
<td>LX2W</td>
<td>AlexNet P FT</td>
<td>4096</td>
<td>76.22</td>
</tr>
<tr>
<td>$l_3$</td>
<td>CNN</td>
<td>LX2W</td>
<td>VGG16 I FT</td>
<td>4096</td>
<td>76.78</td>
</tr>
<tr>
<td>$m_3$</td>
<td>CNN</td>
<td>LX2W</td>
<td>VGG16 P FT</td>
<td>4096</td>
<td>87.28</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Camera Type</th>
<th>Mean</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>RJ</td>
<td>22.44</td>
<td>76.83</td>
</tr>
<tr>
<td>LX2P</td>
<td>12.56</td>
<td>78.83</td>
</tr>
<tr>
<td>LX2W</td>
<td>37.06</td>
<td>82.83</td>
</tr>
<tr>
<td>LX3</td>
<td>29.50</td>
<td>76.50</td>
</tr>
</tbody>
</table>

Results: per-camera statistics.
### Discrimination vs Detection (including rejection)

Real systems should implement a rejection option!

<table>
<thead>
<tr>
<th>Method</th>
<th>Dev.</th>
<th>Options</th>
<th>Dim.</th>
<th>with rejection</th>
<th>w/o</th>
</tr>
</thead>
<tbody>
<tr>
<td>([a_2]) GIST</td>
<td>LX2P</td>
<td>—</td>
<td>512</td>
<td>29.44</td>
<td>42.69</td>
</tr>
<tr>
<td>([b_2]) IFV</td>
<td>LX2P</td>
<td>256</td>
<td>40960</td>
<td>17.50</td>
<td>73.63</td>
</tr>
<tr>
<td>([c_2]) IFV</td>
<td>LX2P</td>
<td>256 SE</td>
<td>41984</td>
<td>12.56</td>
<td>74.44</td>
</tr>
<tr>
<td>([d_2]) IFV</td>
<td>LX2P</td>
<td>512</td>
<td>81920</td>
<td>18.50</td>
<td>67.81</td>
</tr>
<tr>
<td>([e_2]) IFV</td>
<td>LX2P</td>
<td>512 SE</td>
<td>83968</td>
<td>18.00</td>
<td>68.44</td>
</tr>
<tr>
<td>([f_2]) CNN</td>
<td>LX2P</td>
<td>AlexNet I</td>
<td>4096</td>
<td>70.06</td>
<td>87.69</td>
</tr>
<tr>
<td>([g_2]) CNN</td>
<td>LX2P</td>
<td>AlexNet P</td>
<td>4096</td>
<td>64.11</td>
<td>83.19</td>
</tr>
<tr>
<td>([h_2]) CNN</td>
<td>LX2P</td>
<td>VGG16 I</td>
<td>4096</td>
<td>67.28</td>
<td>90.00</td>
</tr>
<tr>
<td>([i_2]) CNN</td>
<td>LX2P</td>
<td>VGG16 P</td>
<td>4096</td>
<td>63.33</td>
<td>76.50</td>
</tr>
<tr>
<td>([j_2]) CNN</td>
<td>LX2P</td>
<td>AlexNet I FT</td>
<td>4096</td>
<td>74.83</td>
<td>84.75</td>
</tr>
<tr>
<td>([k_2]) CNN</td>
<td>LX2P</td>
<td>AlexNet P FT</td>
<td>4096</td>
<td>69.94</td>
<td>81.38</td>
</tr>
<tr>
<td>([l_2]) CNN</td>
<td>LX2P</td>
<td>VGG16 I FT</td>
<td>4096</td>
<td>68.28</td>
<td>88.06</td>
</tr>
<tr>
<td>([m_2]) CNN</td>
<td>LX2P</td>
<td>VGG16 P FT</td>
<td>4096</td>
<td>80.06</td>
<td>88.94</td>
</tr>
<tr>
<td>([a_3]) GIST</td>
<td>LX2W</td>
<td>—</td>
<td>512</td>
<td>39.83</td>
<td>51.75</td>
</tr>
<tr>
<td>([b_3]) IFV</td>
<td>LX2W</td>
<td>256</td>
<td>40960</td>
<td>37.50</td>
<td>73.94</td>
</tr>
<tr>
<td>([c_3]) IFV</td>
<td>LX2W</td>
<td>256 SE</td>
<td>41984</td>
<td>42.83</td>
<td>74.06</td>
</tr>
<tr>
<td>([d_3]) IFV</td>
<td>LX2W</td>
<td>512</td>
<td>81920</td>
<td>39.50</td>
<td>69.69</td>
</tr>
<tr>
<td>([e_3]) IFV</td>
<td>LX2W</td>
<td>512 SE</td>
<td>83968</td>
<td>37.06</td>
<td>68.94</td>
</tr>
<tr>
<td>([f_3]) CNN</td>
<td>LX2W</td>
<td>AlexNet I</td>
<td>4096</td>
<td>75.22</td>
<td>90.31</td>
</tr>
<tr>
<td>([g_3]) CNN</td>
<td>LX2W</td>
<td>AlexNet P</td>
<td>4096</td>
<td>73.89</td>
<td>90.75</td>
</tr>
<tr>
<td>([h_3]) CNN</td>
<td>LX2W</td>
<td>VGG16 I</td>
<td>4096</td>
<td>70.89</td>
<td>90.06</td>
</tr>
<tr>
<td>([i_3]) CNN</td>
<td>LX2W</td>
<td>VGG16 P</td>
<td>4096</td>
<td>81.67</td>
<td>95.44</td>
</tr>
<tr>
<td>([j_3]) CNN</td>
<td>LX2W</td>
<td>AlexNet I FT</td>
<td>4096</td>
<td>73.89</td>
<td>83.19</td>
</tr>
<tr>
<td>([k_3]) CNN</td>
<td>LX2W</td>
<td>AlexNet P FT</td>
<td>4096</td>
<td>76.22</td>
<td>86.94</td>
</tr>
<tr>
<td>([l_3]) CNN</td>
<td>LX2W</td>
<td>VGG16 I FT</td>
<td>4096</td>
<td>76.78</td>
<td>94.81</td>
</tr>
<tr>
<td>([m_3]) CNN</td>
<td>LX2W</td>
<td>VGG16 P FT</td>
<td>4096</td>
<td>87.28</td>
<td>94.88</td>
</tr>
</tbody>
</table>

Shortcomings

1. The rejection option is parametric (we need to learn a threshold);
2. We are not really enforcing temporal coherence (predictions may flicker).

a real system should be able to segment the video into coherent shots
Recognizing Personal Locations (revisited)

**GOAL:** Segmenting egocentric videos into coherent shots related to personal locations.

We collected a dataset of egocentric videos containing 10 personal locations of interest:

- **Training set**: 10-seconds videos (one for each location);
- **Validation set**: one medium-length (5 minutes) video for each location;
- **Test set**: 10 sequences containing transitions between positive locations and negative ones

**Method Overview**

- **1. Discrimination**: Estimation of $P(y_i | I_i, y_i \neq 0)$
- **2. Negative Rejection**: Estimation of $P(y_i | I_i)$ (to enforce temporal coherence)
- **3. Sequential Modelling**: Application of HMM

> arg max $P(y_i = j | I_i)$

> arg max $P(L | V)$

-> generation of segmentation $S$

Method - Discrimination

user-defined set of locations (positives only)

input video

we still want to find this \(-\rightarrow\)

since there are no negatives in the training set, we know that \(y_i \neq 0\)

\[ P(y_i|I_i) \rightarrow P(y_i|I_i, y_i \neq 0) \]

Method - Negative Rejection

How to estimate $P(y_i = 0|I_i)$?

$$Y = \{y_1^*, y_2^*, y_3^*, \ldots, y_K^*\}$$

$$P(y_i = 0|I_i) = 1 - \frac{\sum_j 1(y_j^* = \text{mode}(Y))}{K}$$

variation ratio

Method - Merging Predictions

\[
P(y_i | I_i) = \begin{cases} 
P(y_i = 0 | I_i) & \text{if } y_i = 0 \\ 
P(y_i \neq 0 | I_i) \cdot P(y_i | I_i, y_i \neq 0) & \text{otherwise} \end{cases}
\]

Method - Sequential Modeling

per-frame predictions

HMM

段落化 (Viterbi)

ad hoc "almost identity" matrix

Results

We compare with:

- **SIFT** feature matching to recognize location;
- Open Set Deep Networks (**OSDN**) to reject negatives [1];
- Cascade SVM (**CSVM**) previously investigated [2];
- Entropy Based Rejection (**EBR**) previously investigated [3];
- Negative Trained Network (**NTN**) with one extra class.


Frame Based vs Segment Based Measures

Frame Based

Ground Truth

Segment Based

set overlap threshold (e.g., t=0.5)

choose best match

overlap less than 50%
Results (Segment-Based $F_1$ score)

We compare with:

- **SIFT** feature matching to recognize location;
- Open Set Deep Networks (**OSDN**) to reject negatives [1];
- Cascade SVM (**CSVM**) previously investigated [2];
- Entropy Based Rejection (**EBR**) previously investigated [3];
- Negative Trained Network (**NTN**) with one extra class.

<table>
<thead>
<tr>
<th>Method</th>
<th>$mASF_1$</th>
<th>Car</th>
<th>CVM</th>
<th>Garage</th>
<th>KT</th>
<th>LO</th>
<th>Office</th>
<th>Piano</th>
<th>Sink</th>
<th>Studio</th>
<th>LR</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIFT</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>OSDN [55]</td>
<td>0.47</td>
<td>0.40</td>
<td>0.87</td>
<td>0.54</td>
<td>0.00</td>
<td>0.32</td>
<td>0.08</td>
<td>0.99</td>
<td>0.16</td>
<td>0.99</td>
<td>0.66</td>
<td>0.17</td>
</tr>
<tr>
<td>CSVM [19]</td>
<td>0.49</td>
<td>0.19</td>
<td>0.35</td>
<td>0.44</td>
<td>0.65</td>
<td>0.95</td>
<td>0.57</td>
<td>0.43</td>
<td>0.51</td>
<td>0.48</td>
<td>0.67</td>
<td>0.13</td>
</tr>
<tr>
<td>EBR [16]</td>
<td>0.58</td>
<td>0.36</td>
<td>0.86</td>
<td>0.62</td>
<td>0.54</td>
<td>0.07</td>
<td>0.30</td>
<td>0.84</td>
<td>0.68</td>
<td>0.98</td>
<td>0.97</td>
<td>0.22</td>
</tr>
<tr>
<td>NTN</td>
<td>0.84</td>
<td>0.91</td>
<td>0.93</td>
<td>0.98</td>
<td>0.99</td>
<td>0.49</td>
<td>0.79</td>
<td>0.99</td>
<td>0.89</td>
<td>0.83</td>
<td>0.98</td>
<td>0.49</td>
</tr>
<tr>
<td>Proposed</td>
<td>0.89</td>
<td>0.87</td>
<td>0.94</td>
<td>0.96</td>
<td>0.99</td>
<td>0.96</td>
<td>0.76</td>
<td>0.98</td>
<td>0.83</td>
<td>0.98</td>
<td>0.98</td>
<td>0.54</td>
</tr>
</tbody>
</table>
Qualitative Results

Color-coded segmentation results for qualitative assessment.

Demo Video

http://iplab.dmi.unict.it/PersonalLocationSegmentation/

Personal-Location-Based Temporal Segmentation of Egocentric Video for Lifelogging Applications
A. Furnari, S. Battiato, G. M. Farinella

<table>
<thead>
<tr>
<th>Time Spent at Location</th>
<th>LOC</th>
<th>EST</th>
<th>GT</th>
</tr>
</thead>
<tbody>
<tr>
<td>car</td>
<td>00:00</td>
<td>00:00</td>
<td>00:00</td>
</tr>
<tr>
<td>cvm</td>
<td>00:00</td>
<td>00:00</td>
<td>00:00</td>
</tr>
<tr>
<td>garage</td>
<td>00:00</td>
<td>00:00</td>
<td>00:00</td>
</tr>
<tr>
<td>k. top</td>
<td>00:00</td>
<td>00:00</td>
<td>00:00</td>
</tr>
<tr>
<td>l. office</td>
<td>00:00</td>
<td>00:00</td>
<td>00:00</td>
</tr>
<tr>
<td>office</td>
<td>00:04</td>
<td>00:04</td>
<td>00:04</td>
</tr>
<tr>
<td>piano</td>
<td>00:00</td>
<td>00:00</td>
<td>00:00</td>
</tr>
<tr>
<td>sink</td>
<td>00:00</td>
<td>00:00</td>
<td>00:00</td>
</tr>
<tr>
<td>studio</td>
<td>00:00</td>
<td>00:00</td>
<td>00:00</td>
</tr>
<tr>
<td>l. room</td>
<td>00:00</td>
<td>00:00</td>
<td>00:00</td>
</tr>
<tr>
<td>negative</td>
<td>00:00</td>
<td>00:00</td>
<td>00:00</td>
</tr>
</tbody>
</table>

Estimated Probabilities

<table>
<thead>
<tr>
<th>Estimated Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>car</td>
</tr>
<tr>
<td>coffee v. machine</td>
</tr>
<tr>
<td>garage</td>
</tr>
<tr>
<td>kitchen top</td>
</tr>
<tr>
<td>lab office</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Predicted Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>office</td>
</tr>
</tbody>
</table>

GT Class

<table>
<thead>
<tr>
<th>GT Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>piano</td>
</tr>
<tr>
<td>sink</td>
</tr>
<tr>
<td>studio</td>
</tr>
<tr>
<td>living room</td>
</tr>
<tr>
<td>negative</td>
</tr>
</tbody>
</table>

Location and Object Awareness

The user acquires a 10-seconds video-shot of the personal location

Location Recognition

Living Room
Kitchen
Garage
Negative

Next-Active-Object Prediction
Anticipating the Future

The ability to **anticipate the future** is essential to communicate and interact with each other;

Humans can naturally predict what is going to happen in the short term.

➢ We want the system to be able to anticipate the user’s goals to assist him better.
Can we predict what object the user is going to interact next, before the interaction actually begins?

A. Furnari, S. Battiato, K. Grauman, G. M. Farinella, Next Active Object Prediction from Egocentric Video, under review at Journal of Visual Communication and Image Representation
What defines Next-Active-Objects?

Next-Active-Objects are the next objects the user is going to interact with.

CLASSIC EGO CUES

- Appearance does not change much before the interaction;
- Position alone is not very discriminative;
- Hands are generally visible just before the interaction;

➢ Insight: can we predict which objects are becoming active basing on their trajectories and positions?

A. Furnari, S. Battiato, K. Grauman, G. M. Farinella, Next Active Object Prediction from Egocentric Video, under review at Journal of Visual Communication and Image Representation
Tracks and Trajectories

track: sequence of bounding boxes (one per frame)
trajectory: partial segment of a track

Passive Object Tracks

object is passive

(passive trajectory)

Mixed Object Tracks

object is passive

activation point

object is active

((active trajectory)

A. Furnari, S. Battiato, K. Grauman, G. M. Farinella, Next Active Object Prediction from Egocentric Video, under review at Journal of Visual Communication and Image Representation
Active vs Passive Trajectory Classifier

**Active Trajectory**

**Passive Trajectory**

Random Decision Forest

tree $T$

$P_r(v)$

we need a descriptor

A. Furnari, S. Battiato, K. Grauman, G. M. Farinella, Next Active Object Prediction from Egocentric Video, under review at Journal of Visual Communication and Image Representation
Proposed Trajectory Descriptor

\[ D(T_i) = (xc_1, yc_1, \ldots, xc_h, yc_h, s_1, \ldots, s_h, \Delta xc_2, \Delta yc_2, \ldots, \Delta xc_h, \Delta yc_h, \Delta s_2, \ldots, \Delta s_h) \]

- Position (trajectory)
- Scales
- Differential position (trajectory shape)
- Differential scale («time to contact»)

A. Furnari, S. Battiato, K. Grauman, G. M. Farinella, Next Active Object Prediction from Egocentric Video, under review at Journal of Visual Communication and Image Representation
object track (object detection + tracking)

object is active
object is passive
sliding windows

active/passive trajectory classifier

confidence score

time

A. Furnari, S. Battiato, K. Grauman, G. M. Farinella, Next Active Object Prediction from Egocentric Video, under review at Journal of Visual Communication and Image Representation
Baselines

• Motion Magnitude (moving objects);
• Center Bias;
• Hand Bias;
• Active/Passive Objects (appearance);
• Saliency Based:
  • Fixation Prediction;
  • Salient Object Segmentation;
  • Dynamic Saliency;
• Random.
Results

A. Furnari, S. Battiato, K. Grauman, G. M. Farinella, Next Active Object Prediction from Egocentric Video, under review at Journal of Visual Communication and Image Representation
A. Furnari, S. Battiato, K. Grauman, G. M. Farinella, Next Active Object Prediction from Egocentric Video, under review at Journal of Visual Communication and Image Representation
Conclusion

• FPV allows us to understand what the user is doing;
• FPV can serve as “another pair of eyes” for an assistive robot;
• An assistive system should be aware of location and objects;
  • Wide-Angle/Head-Mounted is better;
  • Continuity of observations can be leveraged;
  • Egomotion can give important insight on “what’s going on”.


Thank You for Your Attention

Antonino Furnari & Giovanni Maria Farinella

http://www.dmi.unict.it/~furnari/
http://iplab.dmi.unict.it/fpv/

PEOPLE

Antonino Furnari  Sebastiano Battiatò  Kristen Grauman  Giovanni Maria Farinella