Robust multi-dimensional motion features for first-person vision activity recognition

Published in: Computer Vision and Image Understanding (CVIU), 2016

Authors: Girmaw Abebe, Andrea Cavallaro, Xavier Parra

Centre for Intelligent Sensing
Queen Mary University of London
Introduction

Third-person vision
Sensing is external

First-person vision
Sensing is ego-centric
Problem definition

- Given first-person videos of ambulatory activities, we develop a robust motion feature in order to recognize the activities.
- Ambulatory activities involve full-body motions.

Third-person view

First-person view
Challenges

• Complex ego-motion
• Motion parallax and blur
  – E.g., Dribble and Sprint
• Local motions
  – E.g., Appearance of people
• Mounting point variations
  – E.g., Chest vs Head
• Limited datasets
Related work: acquisition

• Acquisition device

• Mounting positions
  – Head mounts [Kitani2011, Poleg2016, Poleg2014]
  – Chest [zhang2010,2011]
  – Wrist [Nam2013]

• Preprocessing
  – Data resizing
  – Filtering
Related work: feature extraction

• **Keypoint-based** [zhang2010,2011]

• **Optical flow-based**
  – Magnitude [Kitani2011]
  – Direction [Ryoo2013,2015, Iwashita2014]
  – Frequency domain analysis [Kitani2011]

• **Intra-frame appearance** [Kitani2014,Ryoo2015]
Proposed approach

- Encode significant motion variations
- Generate virtual inertial features
- Multiple validation

• Contributions are highlighted
Grid Features: examples

- Proposed features are shown to discriminate activities.
  - E.g., MDHF: motion-direction histogram feature
Grid Features: more examples

- Frequency feature of motion direction (FTMAF)
  
- Motion magnitude histogram (FTMAF)
Evaluation

• Datasets
  – IAR
  – BAR
  – JPL [Ryoo2013]
  – DogC [Iwashita2014]
### Results on our datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Method</th>
<th>A</th>
<th>P</th>
<th>R</th>
<th>S</th>
<th>F&lt;sub&gt;1&lt;/sub&gt;</th>
<th>KNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP [zhan2015]</td>
<td>92</td>
<td>85</td>
<td>42</td>
<td>99</td>
<td>56</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>IAR</td>
<td>MRGF [zhang2011]</td>
<td>97</td>
<td>90</td>
<td>87</td>
<td>98</td>
<td>88</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>MBH [Kitani2011]</td>
<td>91</td>
<td>62</td>
<td>68</td>
<td>94</td>
<td>65</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td><strong>RMF</strong> (Proposed)</td>
<td><strong>97</strong></td>
<td><strong>91</strong></td>
<td><strong>85</strong></td>
<td><strong>99</strong></td>
<td><strong>88</strong></td>
<td><strong>78</strong></td>
</tr>
<tr>
<td>AP [zhan2015]</td>
<td>90</td>
<td>24</td>
<td>14</td>
<td>97</td>
<td>18</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>BAR</td>
<td>MRGF [zhang2011]</td>
<td>89</td>
<td>35</td>
<td>39</td>
<td>93</td>
<td>37</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>MBH [Kitani2011]</td>
<td>95</td>
<td>63</td>
<td>67</td>
<td>97</td>
<td>64</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td><strong>RMF</strong> (Proposed)</td>
<td><strong>98</strong></td>
<td><strong>81</strong></td>
<td><strong>79</strong></td>
<td><strong>99</strong></td>
<td><strong>80</strong></td>
<td><strong>78</strong></td>
</tr>
</tbody>
</table>

**Key:**
- A: Accuracy
- P: Precision
- R: Recall
- S: Specificity
- F<sub>1</sub>: F-score
- KNN: Recall output of KNN classifier
### Results on public datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Method</th>
<th>A</th>
<th>P</th>
<th>R</th>
<th>S</th>
<th>F₁</th>
<th>KNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>JPL</td>
<td>AP [zhan2015]</td>
<td>76</td>
<td>5</td>
<td>16</td>
<td>86</td>
<td>7</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>MRGF [zhang2011]</td>
<td>85</td>
<td>55</td>
<td>72</td>
<td>87</td>
<td>62</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>MBH [Kitani2011]</td>
<td>87</td>
<td>66</td>
<td>53</td>
<td>92</td>
<td>59</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>RMF (Proposed)</td>
<td><strong>96</strong></td>
<td><strong>87</strong></td>
<td><strong>85</strong></td>
<td><strong>97</strong></td>
<td><strong>86</strong></td>
<td><strong>82</strong></td>
</tr>
<tr>
<td>DogC</td>
<td>AP [zhan2015]</td>
<td>87</td>
<td>39</td>
<td>30</td>
<td>92</td>
<td>34</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>MRGF [zhang2011]</td>
<td>88</td>
<td>39</td>
<td>39</td>
<td>94</td>
<td>39</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>MBH [Kitani2011]</td>
<td>86</td>
<td>38</td>
<td>27</td>
<td>92</td>
<td>32</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>RMF (Proposed)</td>
<td><strong>92</strong></td>
<td><strong>62</strong></td>
<td><strong>59</strong></td>
<td><strong>96</strong></td>
<td><strong>61</strong></td>
<td><strong>58</strong></td>
</tr>
</tbody>
</table>

Key:
- A: Accuracy
- P: Precision
- R: Recall
- S: Specificity
- F₁: F-score
- KNN: Recall output of KNN classifier
Discussion: features
Summary

• Multi-dimensional motion features encode direction, magnitude and dynamics
• Inertial features generated from video proven to be useful
• Collection of new publicly available datasets
• **Limitations**: small datasets, mounting positions (self-occlusions)
Thank you!
Questions?