Person Re-Identification by Efficient Impostor-based Metric Learning

Martin Hirzer, Peter M. Roth and Horst Bischof

Institute for Computer Graphics and Vision
Graz University of Technology
Problem Statement (1)

- Recognize a person across a network of non-overlapping cameras
- Two camera views (persons already detected/tracked):
  - Mark a person in one view (probe image)
  - Find the same person in another video (gallery images)
Related Work (1)

- **Descriptive methods**
  - Distinctive but stable feature representation
  - Spatiotemporal appearance [Gheissari et al. 2006]
  - SDALF [Farenzena et al. 2010]
  - PS [Cheng et al. 2011]
  - ...

- **Discriminative methods**
  - Learn a discriminative model
  - ELF [Gray and Tao 2008]
  - PLS [Schwartz and Davis 2009]
  - ERSVM [Prosser et al. 2010]
  - ...
Related Work (2)

- **Metric learning**
  - Midway between descriptive and discriminative methods
  - Implicit learning of camera transition → suitable for real world scenarios
  - Mahalanobis distance optimal for kNN [Dikmen et al. 2010]
  - Probabilistic metric learning [Zheng et al. 2011]
Mahalanobis Metric Learning

- Mahalanobis distance learning:
  \[ d_M(x_i, x_j) = (x_i - x_j) \mathbf{M} (x_i - x_j) \]

  - Estimate \( \mathbf{M} \) to rescale data
  - Setting \( \mathbf{M} \) to \( \Sigma^{-1} \) : Mahalanobis metric

- Related work:
  - Large Margin Nearest Neighbor (LMNN) [Weinberger et al. 2006]
  - Information-Theoretic Metric Learning (ITML) [Davis et al. 2007]
  - Linear Discriminant Metric Learning (LDML) [Guillaumin et al. 2009]
Learning Metrics from Pairs (1)

- Multi-class problem: ill posed due to high number of classes
- Single-shot re-id task: training data is given via pairs

- Reduce the multi-class problem to a binary problem:
  - Similar pairs: two samples show the same person
  - Different pairs: two samples show different persons

\[
S = \{(x_i, x_j) | y(x_i) = y(x_j)\}
\]
\[
D = \{(x_i, x_j) | y(x_i) \neq y(x_j)\}
\]
Learning Metrics from Pairs (2)

- Exploit discriminative information:
  - Similar pairs should have small distance
  - Dissimilar pairs should have large distance

\[
\mathcal{L}(L) = \sum_{(i,j) \in S} \|L(x_i - x_j)\|^2 - \sum_{(i,j) \in D} \|L(x_i - x_j)\|^2
\]

- Overfitting
  - Biased by dissimilar pairs
Impostors (1)

- Avoid overfitting, focus on the important samples
- Adopt impostor idea of LMNN

\[ I_{(i,j)} = \{(x_i, x_l) | \|x_i - x_l\|^2 \leq \|x_i - x_j\|^2\} \]
Impostors (2)

- New objective function:

\[
\tilde{\mathcal{L}}(L) = \sum_{(i,j) \in S} \| L(x_i - x_j) \|^2 - \sum_{(i,l) \in I} \| L w_l (x_i - x_l) \|^2
\]

- New regularized optimization problem:

\[
\min \tilde{\mathcal{L}}(L) \\
\text{s.t.} \\
L L^\top = I
\]
Efficient Solution

- Define Lagrange function:

\[ \tilde{\mathcal{L}}(L) + \lambda (LL^\top - I) \]

- Setting derivative to zero yields an eigenproblem:

\[ (\Sigma_S - \Sigma_I)L = \Lambda L \]

\[ \Sigma_S = \sum_{(i,j) \in S} (x_i - x_j)(x_i - x_j)^\top \]

\[ \Sigma_I = \sum_{(i,l) \in I} w_{il}(x_i - x_l)(x_i - x_l)^\top \]
Person Re-Identification System

Feature Extraction

Metric Learning

Classification
Experimental Results

- Compare to state-of-the-art
- Compare to other metric learning approaches
- Show generality via multiple different datasets:
  - VIPeR
  - ETHZ
  - PRID 2011
  - PRID 450S
Results - VIPeR (1)

![Graph showing matching rate vs rank for CMC with Proposed and Euclidean methods.](image-url)
<table>
<thead>
<tr>
<th>Method</th>
<th>$r = 1$</th>
<th>10</th>
<th>20</th>
<th>50</th>
<th>100</th>
<th>$t_{train}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>22</td>
<td>63</td>
<td>79</td>
<td>93</td>
<td>99</td>
<td>0.3 sec</td>
</tr>
<tr>
<td>LMNN</td>
<td>17</td>
<td>54</td>
<td>69</td>
<td>87</td>
<td>96</td>
<td>2 min</td>
</tr>
<tr>
<td>LMNN-R</td>
<td>13</td>
<td>50</td>
<td>65</td>
<td>86</td>
<td>95</td>
<td>45 min</td>
</tr>
<tr>
<td>ITML</td>
<td>13</td>
<td>54</td>
<td>73</td>
<td>91</td>
<td>98</td>
<td>25 sec</td>
</tr>
<tr>
<td>LDML</td>
<td>6</td>
<td>24</td>
<td>35</td>
<td>54</td>
<td>72</td>
<td>0.8 sec</td>
</tr>
<tr>
<td>LDA</td>
<td>7</td>
<td>25</td>
<td>37</td>
<td>61</td>
<td>79</td>
<td>0.1 sec</td>
</tr>
<tr>
<td>Euclidean</td>
<td>6</td>
<td>24</td>
<td>34</td>
<td>55</td>
<td>73</td>
<td>-</td>
</tr>
<tr>
<td>ELF</td>
<td>12</td>
<td>43</td>
<td>60</td>
<td>81</td>
<td>93</td>
<td>5 hours</td>
</tr>
<tr>
<td>SDALF</td>
<td>20</td>
<td>50</td>
<td>65</td>
<td>85</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ERSVM</td>
<td>13</td>
<td>50</td>
<td>67</td>
<td>85</td>
<td>94</td>
<td>13 min</td>
</tr>
<tr>
<td>DDC</td>
<td>19</td>
<td>52</td>
<td>65</td>
<td>80</td>
<td>91</td>
<td>-</td>
</tr>
<tr>
<td>PS</td>
<td>22</td>
<td>57</td>
<td>71</td>
<td>87</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PRDC</td>
<td>16</td>
<td>54</td>
<td>70</td>
<td>87</td>
<td>97</td>
<td>15 min</td>
</tr>
</tbody>
</table>
# Results - ETHZ

<table>
<thead>
<tr>
<th>Method</th>
<th>$r = 1$</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SEQ. #1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proposed</td>
<td>78</td>
<td>84</td>
<td>87</td>
<td>89</td>
<td>90</td>
<td>91</td>
<td>91</td>
</tr>
<tr>
<td>SDALF</td>
<td>65</td>
<td>73</td>
<td>77</td>
<td>79</td>
<td>81</td>
<td>82</td>
<td>84</td>
</tr>
<tr>
<td>PLS</td>
<td>79</td>
<td>85</td>
<td>86</td>
<td>87</td>
<td>88</td>
<td>89</td>
<td>90</td>
</tr>
<tr>
<td><strong>SEQ. #2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proposed</td>
<td>74</td>
<td>81</td>
<td>84</td>
<td>87</td>
<td>89</td>
<td>91</td>
<td>92</td>
</tr>
<tr>
<td>SDALF</td>
<td>64</td>
<td>74</td>
<td>79</td>
<td>83</td>
<td>85</td>
<td>87</td>
<td>89</td>
</tr>
<tr>
<td>PLS</td>
<td>74</td>
<td>79</td>
<td>81</td>
<td>83</td>
<td>84</td>
<td>85</td>
<td>87</td>
</tr>
<tr>
<td><strong>SEQ. #3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proposed</td>
<td>91</td>
<td>95</td>
<td>97</td>
<td>98</td>
<td>98</td>
<td>98</td>
<td>99</td>
</tr>
<tr>
<td>SDALF</td>
<td>76</td>
<td>83</td>
<td>86</td>
<td>88</td>
<td>90</td>
<td>92</td>
<td>93</td>
</tr>
<tr>
<td>PLS</td>
<td>77</td>
<td>81</td>
<td>82</td>
<td>84</td>
<td>85</td>
<td>87</td>
<td>89</td>
</tr>
</tbody>
</table>
Results – PRID 2011/450S

- PRID 2011

<table>
<thead>
<tr>
<th>Method</th>
<th>$r = 1$</th>
<th>10</th>
<th>20</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>15</td>
<td>38</td>
<td>50</td>
<td>67</td>
<td>80</td>
</tr>
<tr>
<td>Hirzer’11</td>
<td>4</td>
<td>24</td>
<td>37</td>
<td>56</td>
<td>70</td>
</tr>
</tbody>
</table>

- PRID 450S

<table>
<thead>
<tr>
<th>Method</th>
<th>$r = 1$</th>
<th>10</th>
<th>20</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eukl.</td>
<td>3</td>
<td>15</td>
<td>22</td>
<td>40</td>
<td>52</td>
</tr>
<tr>
<td>Proposed</td>
<td>29</td>
<td>62</td>
<td>73</td>
<td>86</td>
<td>96</td>
</tr>
<tr>
<td>Eukl. + Seg.</td>
<td>13</td>
<td>32</td>
<td>41</td>
<td>55</td>
<td>74</td>
</tr>
<tr>
<td>Proposed + Seg.</td>
<td>35</td>
<td>68</td>
<td>77</td>
<td>90</td>
<td>98</td>
</tr>
</tbody>
</table>
Conclusion and Outlook

- Metric learning for person re-identification
- Efficient implementation
- State-of-the-art results
  - Even though using quite simple representation
  - Drastically reduced computational effort
  - Evaluated on different datasets
- More effective metric learner
- More sophisticated representation
- Extend approach for multi-shot task
Thank you for your attention!

PRID 450S Dataset

lrs.icg.tugraz.at