Shape priors and MRF-segmentation

Boris Flach* and Dmitrij Schlesinger

*Czech Technical University in Prague
Center for Machine Perception
Introduction

In praxi we want to:
- segment objects with inhomogeneous appearance
- model the object shape only approximately

Segmentation model which combines:
- Markov Random Field (Segmentation)
- Level-set based shape priors

What is new?
- Effortless treatment of multiple objects through “decoupling” of segmentation and level-set function
- Statistical formulation allows to pose all learning tasks
The Model

Image domain \( R \subset \mathbb{Z}^2 \), edge set \( E \subset R \times R \),

Sample space \( \Omega = \mathcal{X} \times S \) where

- \( \mathcal{X} \) set of all images \( x : R \to F \) (\( F \) – colour space)
- \( S \) set of all segmentations \( s : R \to K \) (\( K \) – set of segment labels)

**Probability model:**

For each \( k \in K \): \( \phi_k(., \mu_k) : \mathbb{R}^2 \to \mathbb{R} \) a level set function and \( \mu_k \) its pose

\[
P(s; \mu) = \frac{1}{Z} \exp \left[ \alpha \sum_{\{rr'\} \in E} \delta(s_r, s_{r'}) + \lambda \sum_{r \in R} \sum_{k \in K} \delta(s_r, k) \phi_k(r, \mu_k) \right],
\]

\[
P(x|s) = \prod_{r \in R} q(x_r|s_r)
\]
The Model

Comments:
- Model without lateral interaction (Killian et. al. 2004)
- CRF models (Kumar et. al. 2006) have restrictions with respect to learning

Remark:
In what follows, we consider for simplicity $K = \{0, 1\}$ and $\phi_0 \equiv 0, \phi_1 = \phi(., \mu)$. 
Recognition

A, Pose estimation:

We argue for

\[ \mu^* = \arg \max_{\mu} P(x; \mu) = \arg \max_{\mu} \sum_{s \in S} P(x, s; \mu) \]

The function is not concave in \( \mu \).

Applying EM-scheme gives

\[ \begin{align*}
E: & \quad \beta^{(n)}(s_r) = P(s_r|x; \mu^{(n)}) \\
M: & \quad \lambda \sum_{r \in R} \beta^{(n)}(s_r = 1) \phi(r, \mu) - \log Z(\mu) \to \max_{\mu}
\end{align*} \]

The objective \( L(\mu) \) can not be calculated, but interestingly, its gradient

\[ \nabla_{\mu} L = \lambda \sum_{r \in R} \left[ P(s_r = 1 | x; \mu^{(n)}) - P(s_r = 1; \mu) \right] \nabla_{\mu} \phi(r, \mu). \]
Recognition

B. Segmentation:

Loss function: \( \sum_{r \in R} \mathbb{1}\{s_r \neq \hat{s}_r\} \), where \( \hat{s} \) - unknown, but true segmentation.

Minimisation of the risk, i.e. average loss gives

\[
 s^*_r = \arg \max_k P(s_r = k|x; \mu)
\]

Calculation of marginal posteriors \( P(s_r = k|x; \mu) \) is hard. Use Gibbs-Sampler or other approx. methods

Attainable precision of pose estimation and segmentation depends on model parameters \( \alpha, \lambda \) and their interplay.
Recognition

Experiment with synthetic images

Assume a circular shaped object

\[
\phi(r) = \frac{\rho^2 - \|r - z\|^2}{2\rho}
\]

Average Hamming distance estimated - true segmentation:

<table>
<thead>
<tr>
<th>(\lambda) (\alpha)</th>
<th>0</th>
<th>0.7</th>
<th>1.4</th>
<th>1.8</th>
<th>2.1</th>
<th>2.3</th>
<th>3.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>1.e4</td>
<td>1015</td>
<td>107</td>
<td>81</td>
<td>82</td>
<td>86</td>
<td>92</td>
</tr>
<tr>
<td>0.03</td>
<td>1405</td>
<td>189</td>
<td><strong>85</strong></td>
<td>83</td>
<td><strong>82</strong></td>
<td>92</td>
<td>95</td>
</tr>
<tr>
<td>0.06</td>
<td>1075</td>
<td>151</td>
<td>86</td>
<td><strong>80</strong></td>
<td>87</td>
<td><strong>83</strong></td>
<td>92</td>
</tr>
<tr>
<td>0.10</td>
<td>647</td>
<td>126</td>
<td>93</td>
<td>86</td>
<td>90</td>
<td>84</td>
<td><strong>89</strong></td>
</tr>
<tr>
<td>0.30</td>
<td>214</td>
<td><strong>125</strong></td>
<td>113</td>
<td>101</td>
<td>104</td>
<td>98</td>
<td>103</td>
</tr>
<tr>
<td>0.60</td>
<td><strong>187</strong></td>
<td>139</td>
<td>130</td>
<td>123</td>
<td>126</td>
<td>129</td>
<td>136</td>
</tr>
<tr>
<td>1.00</td>
<td><strong>183</strong></td>
<td>152</td>
<td>146</td>
<td>145</td>
<td>148</td>
<td>147</td>
<td>148</td>
</tr>
</tbody>
</table>

\(\lambda^*\) | 0.91 | 0.32 | 0.07 | 0.05 | 0.048 | 0.046 | 0.091 |
Maximum Likelihood Principle:

Consider learning examples as events $A_j \subset \Omega = \mathcal{X} \times S$ and a learning sample $T = \{A_1, \ldots, A_\ell\}$.

$$\log P(T; \alpha, \lambda, \bar{q}) = \sum_{j=1}^{\ell} \log \sum_{(x,s) \in A_j} P(x, s; \alpha, \lambda, \bar{q}) \rightarrow \max_{\alpha, \lambda, \bar{q}}$$

EM-scheme:

**E:** \( \beta_j^{(n)}(x, s) = P(x, s|A_j; m^{(n)}) \)

**M:** \( \sum_j \sum_{(x, s) \in A_j} \beta_j^{(n)}(x, s) \log P(x, s; m) \rightarrow \max_m \)

where $m = (\alpha, \lambda, \bar{q})$.

Substituting the model gives:
New appearance characteristics:

\[ q^{(n+1)}(f | k) \sim \sum_j \sum_{r \in R_j(f)} \beta_j^{(n)}(s_r = k) , \]  

(1)

Gradient with respect to \( \alpha \) and \( \lambda \)

\[
\nabla_\alpha L(\alpha, \lambda, \bar{q}) = \sum_j \sum_{\{rr'\} \in E} \left[ \beta_j^{(n)}(s_r = s_{r'}) - P(s_r = s_{r'}; \alpha, \lambda, \bar{q}) \right]
\]

\[
\nabla_\lambda L(\alpha, \lambda, \bar{q}) = \sum_j \sum_{r \in R} \sum_{k \in K} \left[ \beta_j^{(n)}(s_r = k) - P(s_r = k; \alpha, \lambda, \bar{q}) \right] \phi_k(r, \mu_k) .
\]

We tested learning of \( \lambda \) for certain fixed \( \alpha \) values.
Results
Conclusions/Open Questions

The model

- allows to integrate (multiple) shape priors and appearance models into MRF-segmentation
- allows concise formulation of recognition and learning tasks
- improves segmentation even using relatively simple shape models

Open questions:

- More complex, statistical shape models (pose no longer a parameter).
- How to deal with wrong pose initialisations?
- How to model occlusions?
- How to model shape priors in a distributed way?
Outlook: Distributed Shape Priors

Actually we study:

- “Translational invariant” MRF-s of second order (as for texture modelling)
- Model Gibbs potentials for all edges bounded by a certain “length”
- Labels denote e.g. object parts

Preliminary results: