Detecting Gestures in Medieval Images

Joseph Schlecht  Bernd Carqué  Björn Ommer

**Introduction**

Gestures in medieval manuscripts symbolize communication.

- Historical origin and usage of gestures not well understood.
- Sachsenspiegel is one of the earliest books of German law.
  - Elke von Reggow, ca. 1220-1235
  - Heidelberg, Dresden, Wolfenbüttel and Oldenburg

Position of arms, hands and fingers indicate communication.

Gestures interpreted semantically by legal historians.

- Voting with correlation response over scale and orientation.
- Suggests a systematic comparison using automated gesture detection.
  - Bonus: Characterize artistic variation across workshops and draftsmen.

**Approach**

Capture gesture appearance variation with a set of learned templates.

- Cross-correlate image $f$ with template $t_i$ and normalize.
- Compute principle components $t = UVW^*$.
- Take $t_i$ with largest projection on first $N_p$ principle components of $U$.
- SVM trained in two stages: one-vs-rest, then find false-positives and retrain.

**Detection**

Given a set of templates capturing gesture variation $t_1, \ldots, t_N$.

Cross-correlate image $f$ with template $t_i$ and normalize.

$$h_i(x,y) = \sum_{u,v} \frac{f(x-u,y-v)t_i(u,v)}{M_i}$$

Variance of image under template expensive, use integral image.

$$s_{xy}^2 = \frac{1}{M_i} \sum_{u,v} \left( f(x-u,y-v) - \bar{f}(x,y) \right)^2$$

Collect correlation response over position, angle and scale.

$$H(x,y,\sigma,\phi) = \sum_{i=1}^{N_s} h_i(x,y;\Phi(\lambda_i))$$

- Apply HoG and SVM for verification.

**Learning**

Ground-truth gestures labeled with oriented bounding boxes.

- Rotate and scale normalize each gesture type; convolve with Laplacian of Gaussian.
- Compute principle components $t = UVW^*$.
- Take $t_i$ with largest projection on first $N_p$ principle components of $U$.
- SVM trained in two stages: one-vs-rest, then find false-positives and retrain.

**Results**

Ground-truth comprises 280 scenes from Heidelberg Sachsenspiegel.

- 347 labeled gestures, each about 200x100 pixels.

Learn 10 templates using 10 principle components.

Evaluate detector with leave-one-out strategy.

- Successful recovery of orientation statistics.

75% of all gestures detected at $< 1$ FPPI.

**Future work**

- Detect gestures across other Sachsenspiegel manuscripts.
- Scenes have concordance, utilize context for alignment.
- Identify changes in gestures.