Compact Part-Based Image Representations
Marc Goessling, Yali Amit
Department of Statistics

Introduction
Learning compact, interpretable part-based image representations is still an unsolved task. We review various existing composition rules for binary data and introduce the max-minus-min rule. We also propose a novel sequential initialization procedure based on a process of oversimplification and correction. The experiments show that our approach leads to very intuitive models.

Composition rules
The (binary) image data \( l \) is modeled through a Bernoulli distribution \( P(l | \mu) \) where the global template \( \mu(x) = \gamma(\mu_1(x), \ldots, \mu_K(x)) \) is a composition of part templates \( \mu_k \) which are defined on the entire image grid. Different composition rules \( \gamma : [0, 1]^K \rightarrow [0, 1] \) can be considered [1-4]. We propose to use the max-minus-min rule (where \( q \) specifies ‘no opinion’) \[
\gamma(p_1, \ldots, p_K) = q + (\max_k p_k - q) - (\min_k p_k - q),
\]
which reduces redundancy (since only the most extreme template votes) and encourages vote abstention (because opposing opinions are penalized strongly).

Geometric component
The spatial arrangement of the parts is modeled as a joint Gaussian distribution on locations and orientations.

Inference: Likelihood matching pursuit
Given current templates \( \mu_k \) the task is to find the part configuration \( (\mu_1, \ldots, \mu_K) \) which maximizes the likelihood of the image data \( l \):
\[
\text{set } \mu(x) = q
\]
REPEAT
\[
\begin{align*}
& \text{find } k^*, t^* = \arg\max_k \ P(l | \gamma(\mu(x), \mu_k(t(x)))) \\
& \text{update } \mu(x) = \gamma(\mu(x), \mu_{k^*}(t^*)) \\
& \text{UNTIL no improvement is possible anymore}
\end{align*}
\]

EM Learning
Define the parts with the most extreme opinion for pixel \( x \):
\[
k^*(x) = \arg\max_k \mu_k(x), \quad t^*(x) = \arg\min_k \mu_k(x)
\]
In the M-step we update the parts by computing
\[
\mu_k(x) = \frac{\sum_k \mathbb{1}(k = k^*(x) \text{ or } k = t^*(x)) \phi^2_k(l(x))}{\sum_k \mathbb{1}(k = k^*(x))}
\]
which is simply the average of all (back-transformed) images for which the part was responsible.

Sequential initialization
Good initializations are crucial because the learning problem is non-convex. We start with over-simplified models which try to explain the data using only very few parts. These models are then ‘corrected’ by appending residual images (difference between a training example and the model explanation) as additional parts.

Part transformations
Explicitly modeling shifts and rotations allows to share parameters among all transformed versions \( \mu_{k,t} = \Phi_t(\mu_k) \) of the part template \( \mu_k \).

Handwritten letters
We train models with up to 4 parts on the letter classes from the TiCC handwritten characters dataset [5].

Figure 1: Top: Asymmetric and symmetric composition rules, as a function of \( p_k \) for \( p_k \geq 0.7 \). Bottom: Compositions of two parts using the different rules (dark means higher probability). The probabilities in the first template are 0.5 and 0.7, the probabilities in the second template are 0.7 and 0.01.

Figure 2: Learning a part model for the letter T. 1st row: The 10 examples used for training. 2nd & 3rd row: Online learning of two parts. Shown are the two templates at step \( t = 1, \ldots, 10 \) (blue corresponding to 0, yellow corresponding to 1). 4th row: Sampled part configurations using a multivariate Gaussian distribution on the spatial arrangement of the parts.

References

http://galton.uchicago.edu/~goessling/