Supplementary Material

Factored Shapes and Appearances for Parts-based Object Understanding

S. M. Ali Eslami
Christopher K. I. Williams

School of Informatics,
University of Edinburgh,
United Kingdom

July 27, 2011

A Results on synthetic data

Consider a synthetic dataset of two bars of variable length such as the one shown in Fig. 1. Notice that in each image the two bars are synchronised (*i.e.* their lengths are equal), and that the red bar always occludes the blue bar. We train a global FSA model ($L = 2$, $H = 1$) on this dataset. The data is trivial to segment with appearance cues alone, but we focus on the way in which FSA learns to model the shapes of the two bars.

First we plot the learned appearance model upon convergence of the learning algorithm in Fig. 2. In Fig. 3 we plot samples of the learned shape model. Notice how the bars vary in length, are always of the same size, and appear with the correct occlusion ordering. In Fig. 4 we show the way in which the image structure varies as $\mathbf{v}$ moves in 1D space.

We also train a *local* FSA model ($\bar{H} = 1$ per layer), and plot the samples it generates in Fig. 5. The generated bars now appear with different lengths in the same image. The model has *generalised* from the training data. Notice how the blue bar is ragged at its tip. This is to be expected, since the model has never actually 'seen' what the tip should look like in the training data – it has always been occluded by the red bar in that region.

We train the same local FSA model again, but this time on an unsynchronised dataset in which the two bars appear with different lengths in each image. We plot samples from the model in Fig. 6, and in Fig. 7 we show the way in which the model's distribution on image structure varies as $\mathbf{v}$ moves in 2D space.

These results demonstrate that FSA can learn about the shape variability observed in the data. The global FSA model faithfully captures the covariance of the two bars, and the local model can be used to impose prior knowledge in order to make it generalise from what it has seen.

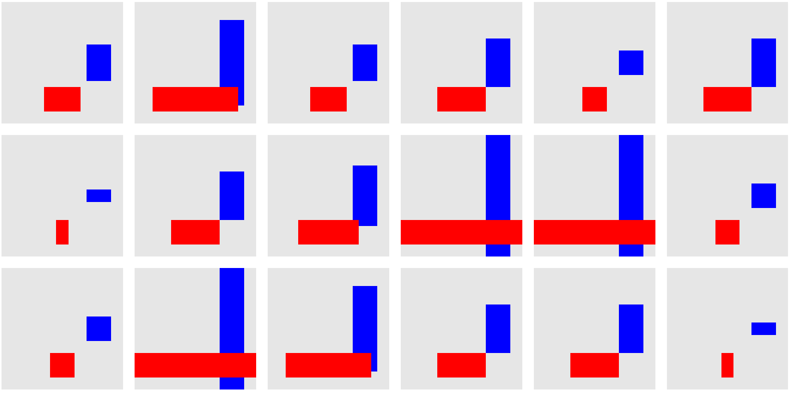

Figure 1: A subset of the training images.

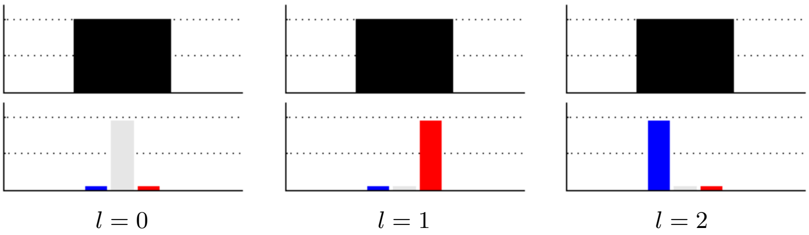

Figure 2: The learned appearance model.

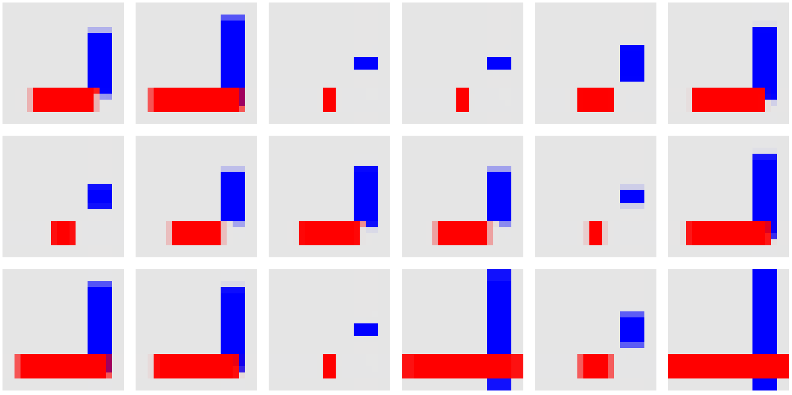

Figure 3: Random samples from the learned model. Here we consider a global model with $H = 1$. The two bars are always of the same length.

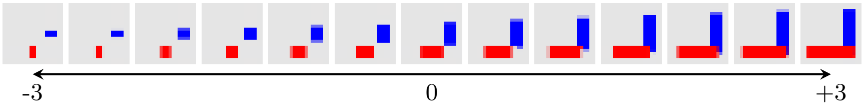

Figure 4: The image structure varies as $\mathbf{v}$ moves in 1D space. Here we consider a global model with $H = 1$. The two bars are always of the same length.

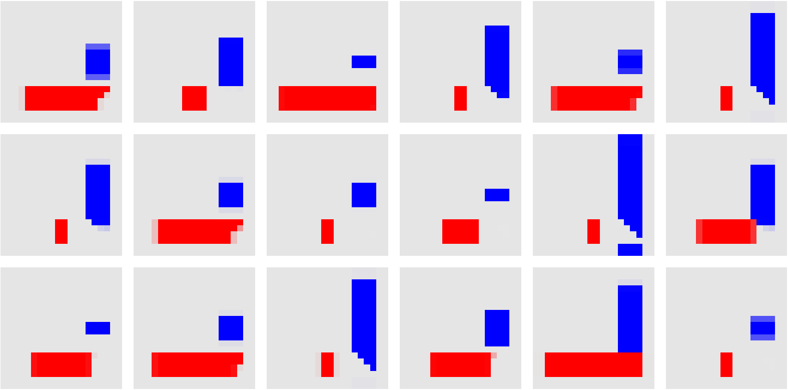

Figure 5: Random samples from the learned model. Here we consider a local model with $\bar{H} = 1$. Synchronised training data. The two bars are of varying lengths. The blue bar is at times ragged.

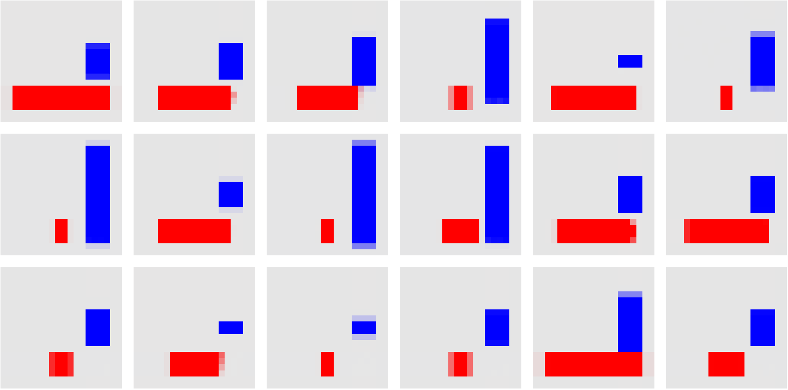

Figure 6: Random samples from the learned model. Here we consider a local model with $\bar{H} = 1$. Unsynchronised training data. The blue bar is no longer ragged.

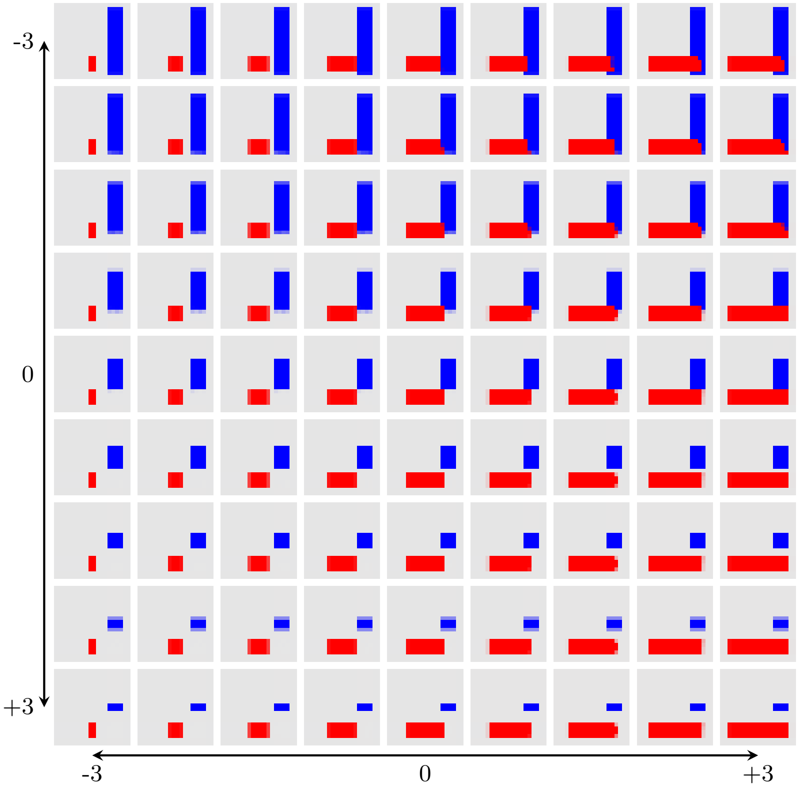

Figure 7: The image structure varies as $\mathbf{v}$ moves in 2D space. Here we consider a local model with $\bar{H} = 1$. Unsynchronised training data.

B Supervised training on Cars dataset

Fig. 8 illustrates the shape model learned by FSA when trained in a supervised manner on the Cars dataset. Qualitatively, one can see that the unsupervised model's performance is comparable to that of the supervised model. In the top-right corner, one can see the prototypical *SUV* shape. In the bottom-right we see the *saloon* shape, in the bottom-left we see the *hatchback* shape, and the *convertible* shape can be seen in the top-left.

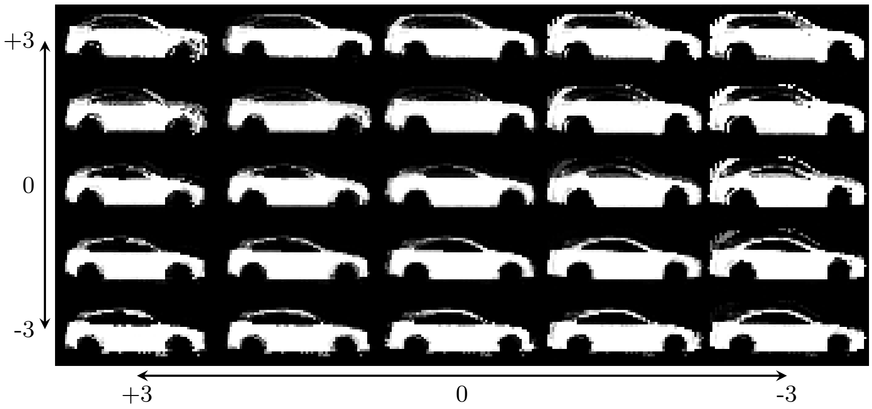

Figure 8: A plot of the car body's mask for a grid of $\mathbf{v}$ values in 2D latent space.