V. Cheung, N. Jojic, and D. Samaras. Capturing long-range correlations with patch models. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2007.

The videos have been encoded in Quicktime format using H.264 compression.

Fig 1 - Car epitome.mov

This video illustrates the learning procedure that leads to the estimation of the epitome for the three cars shown in Fig. 1. The learning took 5 iterations of mapping estimation and epitome re-estimation as discussed in Sections 2 and 3. The mapping estimation is itself an iterative process (10), performed 7 times each iteration, illustrated in the phases of the video marked “Mapping estimation”. Patches are moved according to their mapping T to the epitome. Once the mapping stops changing, the patches are then averaged to create a new epitome, symbolically shown in the parts of the video marked as “Epitome estimation” as the flight of the patches to the previous epitome estimate, after which the new estimate, obtained according to (5) is shown. The process repeats until the epitome converges, which in this case, was obtained in 5 iterations. The epitome ends up being a merge of the three cars, with both the front and back of the car reaching a compromise between the different shapes of the cars. Also interesting to note is how the epitome merges the three backgrounds. The car images and the epitome all have a resolution of 120x90 pixels, and patches of size 10x10 with 10 random correlation links were used during learning.

Fig 4 - Face mapping.mov

The video illustrates the mapping of three face-like images from Fig. 4 to the face epitome shown on the right in Fig. 3. The mapping is estimated by iterating (10) and is illustrated in the same manner as in the video corresponding to Fig. 1. However, after the mapping, rather than updating the epitome, the epitome is “lit-up” proportional to its usage for that image, as described in Section 5.1. The true face image at the top-left is nicely mapped to a face in the epitome, and the synthetic Cyclops image on the middle-left is mapped to the same face by moving the single eye into the position of the left eye in the epitome, and deforming the rest of the face to compensate for the violation of the constraints g (7). Due to the cost of these violations, the likelihood of the Cyclops image is lower than the likelihood of the true face image. The image of the dog, on the other hand, while having the facial features in a better spatial arrangement, does not match in the appearance, due to the dog’s fur texture, and the best match still requires breaking the optimal spatial configuration. The dog image likelihood is thus penalized both due to the poor patch likelihoods (9) in factors h, and due to the violation of the constraints g in (7), and is the lowest among the likelihoods for the three images. (Likelihoods are not normalized, but can still be used to rank images - see footnote 5). The faces are of size 33x50, with the dog image of size 49x50. The epitome is 96x96 and patches of size 8x8 were used.

Fig 5 - Face relighting.mov

The video illustrates a video of a face undergoing illumination change, which we used as an epitome for the task of relighting a photograph by synthesizing a video volume whose one frame is initialized to the target photograph, and the rest are extrapolated by iterating (10, 15, 16, 17, 14), as described in Section 5.2. In this process, a mapping similar as the one illustrated in the previous two videos, but in 3D (x, y, and illumination angle), is iterated with estimation of the means \nu of the interpolated video. The result is a plausible relighting of the face in the target photo. Note that the synthesis result has sharp shadows not only on the nose, but also the background, shoulder, and neck. The training video of size 105x130 was used to synthesize 28 frames of size 100x125 from the target frame using patches of size 10x10x5 with 30 correlation links.

Fig 6 - Face relighting without patch correlations.mov

This video repeats the same experiment as that in Fig. 5, using the same example video, but with a different target photograph. Two results are shown. The result on the left was synthesized with no constraints on the mappings, and the result on the right uses the flexible patch configurations model. The constraints result in a smoother and more consistent synthesis, especially the further you are away from the initial seed frame. 197 frames of size 130x150 are shown.

Fig 7 - Cloth relighting.mov

This is the same as the video for Fig. 5, but this time, we are relighting a piece of cloth, based on a small sample of how an image of a different drape of the same material changed with illumination. Both the training set and the test image were of size 150x150, from which 74 frames were extrapolated using patches of size 15x15x5 with 50 correlation links.

Fig 8 - Image walkthrough.mov

The same as the videos for Fig. 5 and 7, but this time, the third dimension is time t, and the interpolation of a video volume from one frame leads to simulating walking through a hallway in the photograph. Using patches of size 5x5x3 and 20 random correlation links, 10 frames of resolution 180x120 were synthesized that exhibit a plausible movement of walls, lights, and fixtures as if you were walking through that hallway.