Modeling Object Appearance using Context-Conditioned Component Analysis

Daniyar Turmukhambetov [1], Neill D.F. Campbell [1,2], Simon J.D. Prince [1,3], Jan Kautz [1,4]
[1] University College London, UK. [2] University of Bath, UK. [3] Anthropics Technology, UK. [4] NVIDIA, USA.

Limitations of Existing Subspace Models: Subspace models have been very successful at modeling the appearance of structured image datasets when the visual objects have been aligned in the images (*e.g.*, faces [2, 3]). Even with extensions that allow for global transformations or dense warps [1] of the image, the set of visual objects whose appearance may be modeled by such methods is limited. They are unable to account for visual objects where occlusion leads to changing visibility of different object parts (without a strict layered structure) and where a one-to-one mapping between parts is not preserved. For example bunches of bananas contain different numbers of bananas but each individual banana shares an appearance subspace.

Varying Structure: Visual objects with varying structure cannot be readily expressed using previous methods. By fixed structure, we mean that images consist of a *fixed set* of regions that have certain associated textures. For example, side views of cars in general have *two* regions corresponding to two wheels and these regions are always present in the image. On the other hand, an example of a visual object with a varying structure would be a facade of a building with a *varying number* of windows. In this scenario, the texture of window regions can be shared across images, but each image may have a different number of regions corresponding to windows. Furthermore, datasets of visual objects that have occlusions or significant deformations which can cause self-occlusions or 2D topology changes also remain a challenging input for current methods, *e.g.*, animals under different poses.

Our Work: In this work we remove the image space alignment limitations of existing subspace models by conditioning the models on a shape dependent context that allows for the complex, non-linear structure of the appearance of the visual object to be captured and shared. This allows us to exploit the advantages of subspace appearance models with non-rigid, deformable objects whilst also dealing with complex occlusions and varying numbers of parts. Specifically, we develop a model called Context-Conditioned Component Analysis that approximates images with a linear combination of basis functions, which are shared between all images. The function weights provide a low-dimensional representation of the appearance. The argument of the basis functions is a *context vector* that provides some information about the state of the object at that pixel. For example, it might encode the part of the object at that pixel (door, wall of a house), the local shape of the silhouette of the object, or the distance from some pre-determined keypoint. We derive a generalized EM algorithm for learning the model. We note that our model encompasses both Probabilistic Principal Component Analysis and Active Appearance Model as special cases.

Conclusion: We demonstrate the effectiveness of Context-Conditioned Component Analysis with examples of structured inpainting and appearance transfer on 4 datasets of varying difficulty. Figure 1 shows some results of the appearance transfer, where the appearance of one instance of a cat (*e.g.*, in a specific pose) is transferred to others (*e.g.*, different poses); we note that complex structure is well modeled, for instance the differing colors of the body of the cat in the last column. Figures 2 and 3 show structured inpainting of unobserved test images for horses and elephants, respectively.

[1] Timothy F Cootes, Gareth J Edwards, and Christopher J Taylor. Active appearance models. In *Computer Vision – ECCV 1998*, pages 484–498. Springer, 1998.

[2] Umar Mohammed, Simon J. D. Prince, and Jan Kautz. Visio-lization: generating novel facial images. *ACM Trans. Graph.*, 28(3):57:1–57:8, July 2009. ISSN 0730-0301. doi: 10.1145/1531326.1531363.

[3] Matthew A Turk and Alex P Pentland. Face recognition using eigenfaces. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 586–591. IEEE, 1991.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

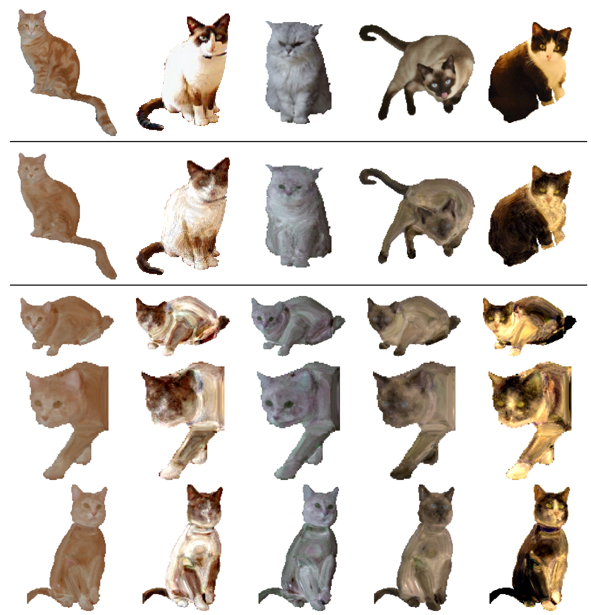

Figure 1: Appearance Transfer Results (Cats). For each column, rows top to bottom: (1) An image of the training set; (2) Reconstruction of the top row image; (3-5) Reconstruction of the image of *another* cat in a different pose using appearance parameters of the top row image.

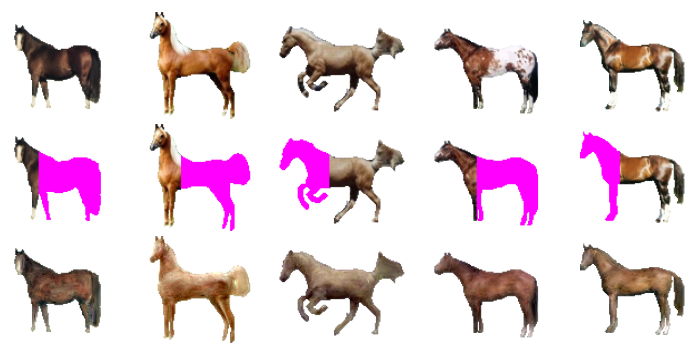

Figure 2: Image Inpainting Results for Horses (Test Set). For each column, rows top to bottom: An image from the test set; Input for inpainting (context vectors of all pixels are known); Inpainted result.

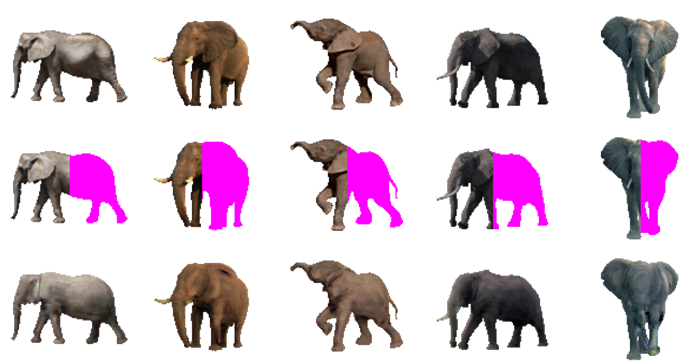

Figure 3: Image Inpainting Results for Elephants (Test Set). For each column, rows top to bottom: An image from the test set; Input for inpainting (context vectors of all pixels are known); Inpainted result.