Photo Clip Art

Zainab AlMeraj

Jean-François Lalonde, Derek Hoiem, Alexei A. Efros, Carsten Rother, John Winn, and Antonio Criminisi. Photo clip art. In *SIGGRAPH '07: ACM SIGGRAPH 2007 papers*, page 3, New York, NY, USA, 2007. ACM

Lalonde et al.'s Photo Clip Art introduces a 3D photo-realistic photo clip art library that allows plucking images of objects into existing photographs by selecting only the appropriate ones based on the corresponding image parameters (i.e. such as lighting). The authors accomplish this in two steps: using data-driven searches, to acquire relevant objects rather than starting with a particular object; interpreting images in 3D, to acquire an understanding of the geometry of the scene as well as it's illumination context. Their algorithm includes novel variations on existing image-based segmentation algorithms, a first attempt at automatically estimating the 3D geometry and illumination of images for interactive use, as well as novel improvements on image blending routines required for smooth final images. Then given a scene incorporated in a neat user interface, a user may query a categorically subdivided database for good objects to insert. The goal is to help the user avoid any manual image composition whilst still achieving impressive photo-realistic results easily.

To allow for interactive data-driven searches, the user must be able to query a pre-computed library of objects. This utilizes a modified database from "LabelMe", a large publicly available database of labeled objects. Over 13,000 objects are classed into categories and also sub-categorized into visually similar subclass of objects for easier access. An inference model successfully estimates camera height and position within a scene given only an approximate height distribution of any one on the object classes. If $\mathbf{y_i}$ is the 3D height of an object with a 2D height of $\mathbf{h_i}$ and a vertical position $\mathbf{v_i}$, and the camera height and horizon position are $\mathbf{y_c}$, $\mathbf{v_0}$ respectively. Then $\mathbf{y_i} = \mathbf{h_c} \times \mathbf{y_c} \div (\mathbf{v_0} - \mathbf{v_i})$. This is then inferred over the entire image dataset. Incident illumination of objects are gathered from approximate environment maps captured for all database images. These are acquired through visualizing rough 3D structures of the scene and looking at the distribution of colour shades within three broad regions, ground, vertical planes and sky, in the image. These parameters are then added onto existing object annotations.

Once an object is selected, and using the gathered knowledge about the position, pitch and height of the camera and light sources, the algorithm then looks for images in the clip art database that were taken from similar positions and with similar pixel heights. Lalonde et al. introduce the use of shape priors for image segmentation along with two other classifications that triumph existing segmentation and image matting approaches. The priors are meant to adjust original database segmented images that lack detailed outline and decent alpha masks. These algorithms are successful at conserving original shapes, recovering full alpha matte's for foreground and background of objects and preventing discolouration when blending to the scene. A final image-based shadow transfer gives objects a sense of sitting on the ground once placed in the scene.

Adding synthetic objects to a scene, finding global illumination, lighting and the geometry of a scene are not new concepts. The novelty of this paper lies in its ability to capture camera properties and illumination conditions from a singe photograph regardless of the location it was taken, and using this information to query objects with the similar properties. The authors emphasize the systems ease of use and applicability to novice users. But initial attempts at object insertion and achieving novel visual content are not as straightforward as one may expect.

This paper contains a number of algorithms and techniques that are utilized throughout the process of scene object insertion. The concepts for each step of the system are well explained and the author's extensions onto the previous methods (which are many in number) are fairly justified. The authors succeed in hiding the complexity of all these algorithms from the user with a neat interface and make the process seem very simple and intuitive. More indication as to how well certain object resolutions look in scene would be beneficial. As the authors indicated, more data does solve many issues relating to object matching reliability. But since lighting is approximated using global matching, some elaboration may have been interesting as to how their local context (area around an object) particularly handles the shadow cases.