Sketch2Photo: Internet Image Montage

David Fagnan

Tao Chen, Ming-Ming Cheng, Ping Tan, Ariel Shamir, and Shi-Min Hu. Sketch2photo: Internet image montage. In *ACM SIGGRAPH ASIA 2009, ACM Transactions on Graphics*, 2009

Although not everyone can draw or sketch a masterpiece, almost everyone can sketch stick figures and objects. Sketch2Photo creates a process to automatically generate a photomontage from such simple hand drawn sketches, where each object is given a text label. Currently, *photomontage* can be done using many commercial applications (eg. Adobe Photoshop) but requires the user to provide suitable images to be composed. The quality of the resulting image then depends on various factors unknown to novice users, such as illumination and texture. This is accomplished by searching online for each object in the sketch via its text label. The authors use several steps in the process including image search, image segmentation and image composition (blending), each of which is a non-trivial problem in general. The focus of the paper involves searching for images via online image databases and filtering these based on how algorithm friendly they are, in order to avoid algorithm weaknesses.

In addition to restricting the background to a landscape, the authors also apply content consistency filtering and uncluttered region filtering. Content consistency filtering groups the candidate images based on content, where it is then assumed that the largest group is the best matching set of images. In particular, the feature considered is a histogram in LUV color space, where cluster distances are computed to the largest cluster and about 10% of the total images are kept. Lastly, the final filter on background images uses segmentation to estimate the number of segments covered by the area enclosed by scene items in the input hand sketch.

The scene items are also filtered and extracted to be more algorithm friendly, using saliency filtering, scene item segmentation, and both contour and content consistency filtering. After ignoring images without clear simple backgrounds, saliency filtering then counts the number of segments surrounding the high-saliency region as a measure of clutter in the image, since the desired object is expected to be central in the image. The high saliency region is then expanded, and a grab-cut algorithm is applied to this region to extract the scene item. Hand-drawn contours are then compared to the contour of the saliency-extracted scene object, and only a fraction of the candidate images are selected for further processing. The final step for scene items, is a content consistency filter similar to that used on background images, but allowing for multiple target clusters.

The paper demonstrates drawbacks of the two major types of seamless image composition techniques (alpha blending and Poisson blending). The authors introduce a combination on these algorithms that improves image composition and avoids some of the weaknesses of each technique. The algorithm involves computing an optimal blending boundary with two groups, one containing pixels suitable for Poisson blending (based on a estimated blending cost), while the other group may suffer from colour and texture differences. The improved blending technique basically performs Poisson blending on the pixels that are found to be suitable, and attempts to do something better by separating the source foreground with a seperate boundary condition and blends the resulting image a second time using alpha blending.

Although all the algorithms used in this paper are well studied and while the authors use mostly existing algorithms, there are two key new ideas in this paper. The first is the overall scheme; using various algorithms to filter a selection of candidate images to those that will perform well with the algorithms required for image segmentation and composition. Through their effective user study, the authors show that the interaction time among novice users is reduced by about a factor of 5-10 when compared to photoshop or drag and drop composition. The work does an excellent job of detailing minor steps required in the process, including interactive refinement which allows the user to highlight problem areas in the image to further correct algorithms.

The second idea is the introduction of the improved blending method in the paper, which although it is impressive, it may have been better suited in a longer report, or a paper devoted solely to this method. As it stands, the authors provide only one successful example of the blending, and do not elaborate on the performance other than to claim that it does not suffer from texture mixing, colour discolouration, or contrasting lighting. However, the authors do an excellent job pointing us to failures in the other steps of the work and offer possible solutions that can be explored in future work. These include, incorrect scene objects, incorrect perspective, and incorrect occlusion. As result natural extensions of the work could include estimating camera pose and exploring further methods of avoiding incorrect scene objects appearing in candidate compositions.