Cubist Rendering with Object Extraction

Elio G. Champenois

University of California, Berkeley

CS294-26 Fall 2015 Final Project
Abstract

The Cubist art style explored capturing a subject or scene from multiple viewpoints in one image. In this project, starting with these multiple viewpoints in the form of several photographs of a subject, a Cubist style rendering is achieved in a multi-step process, guided in part by the work of Collomosse and Hall [IEEE TVCG, Vol. 9, No. 4, October-December 2003], but without the need for user interaction. First, important features in the images are identified as regions of high saliency. Next, a composition is found by adding non-overlapping features from distinct images. Non-salient regions are then added to the composite image, chosen in each region from the appropriate original image. Finally, the image is rendered giving a painted style through block segmentation, warping, shading, and point stroking.

Keywords: Cubism, Rendering, Salient
Cubist Rendering with Object Extraction

A method for nonphotorealistic rendering in a style of Cubist art is presented. The method is inspired by the work of Collomosse and Hall, making use of feature extraction from several images to find a composition that avoids stitching separate viewpoints through important points of the images [1]. Features are found by analyzing the distribution of the salience of the images. In this previous work, the user is asked to pick out and label the features (after an initial salience analysis displays where these features likely are) into several groupings, and a final composition is created, avoiding feature overlap and also attempting to fix the amount of features belonging to a specific feature class the same as in the original images.

In this work, the need for user input is avoided and replaced with automatic feature extraction. This method, though ignorant of what the found features represents, is still able to make reasonable compositions by better analyzing the spatial information of these features in the separate images.

In this paper, this method for feature extraction and image composition will first be presented. The composition is followed by the rendering, which includes several steps: (1) features are warped for a surrealist effect, (2) image segmentation is used for gradient shading, for a brush stroke effect, and (3) an algorithm adapted from the work of Sparavigna and Marazzato is used for a further Impressionist rendering [2].

Feature Extraction

Features are extracted from the original images by making use of the salience of each image, as calculated using a technique adapted from work by Walker et al. [3]. The convolutions
of each image with 15 different filters are calculated. These filters are the horizontal and vertical first and second order Gaussian derivatives, notably $\partial G(x, y; \sigma)/\partial x$, $\partial G(x, y; \sigma)/\partial y$, $\partial^2 G(x, y; \sigma)/\partial x^2$, $\partial^2 G(x, y; \sigma)/\partial y^2$, and $\partial^2 G(x, y; \sigma)/\partial x \partial y$, for three different values of the Gaussian width $\sigma$. In a region with an interesting feature, at least one of these filters respond, due to the different directionalities probed by the different filters and the different scales probe by the different Gaussian widths. The values of all of these convolutions at each pixel create a vector $\vec{x} \in \mathbb{R}^{15}$. For each image, the mean of the vectors at each pixel, $\vec{\mu}$, is subtracted from each of these vectors, and the principal components of the distribution of these vectors are found with eigenvalue analysis, giving the covariance matrix for the distribution, $\vec{V}$. The salience of the analyzed image at a given pixel is then calculated using the squared Mahalanobis distance of the vector at that pixel, defined as:

$$(\vec{x} - \vec{\mu})^T \vec{V}^{-1} (\vec{x} - \vec{\mu})$$

To find candidate features, all of the points above some salience threshold that are also local maxima are selected. Next, adaptive non-maximal suppression (ANMS) is used to find some subset of these candidate features that are both highly salient and distributed evenly about the image. Figure 1 shows a typical output of this salience calculation as well as the found feature points.

The image is then segmented into regions, where each region is associated with one of these feature points through a Voronoi diagram. For each region, we apply a threshold to the salience of the image followed by several binary erosions. Finally, we look for the largest contiguous binary object and define that as the mask for the found feature in that region. This technique is generally successful at picking out small to medium sized features distributed evenly around the image.
Image Composition

Although the features found in one image will be spatially separated, features from different images may overlap. This is accounted for to generate the final image composition. In the previous work, the candidate features were filtered and labeled by the user, and this allowed selecting a feature subset that kept some resemblance to any one input image. Here, we find that similar results are achievable by simply looking for strong features that do not overlap. This is not always the case, but, with slight parameter tuning, has worked for all tested image data sets. The success of this approach is attributed to the fact that the images represent the same subject in a similar position in the photograph, with the viewpoint being the only change. Features from separate images that might have been labeled by the user as belonging to the same feature class will be spatially close to each other. While the previous work addresses the doubling of these features through a stochastic selection process, the spatial filtering method described below will achieve a similar result given input images which follow the aforementioned style.

All of the found features from all of the images are listed and sorted by their mean salience, with features with higher salience appearing first. We begin with an empty composite image and iterate through this list: if the composite image at the location of the current feature is empty, that feature is added to the composition. If there is some overlap between the current feature and the current composite image, the feature is thrown away. Figure 2 shows the result of this method for portrait images.
Superquadratic Warping

In following with the previous work, we apply a similar superquadratic warping to the original image. This superquadratic warping creates a surrealist effect, squaring off the corners of ellipses. For each picture, we look individually at each feature that has been selected for the final composition that originates from that source image. The edge of the feature is fit using least squares with an ellipse (superellipse with $\alpha = 2$), with parameters for the center $(x_c, y_c)$, axis radii $(r_x, r_y)$, and tilt $t$:

$$
\begin{align*}
    x &= \frac{(x_i - x_c) + t(y_i - y_c)}{r_x}, \\
    y &= \frac{(y_i - y_c) - t(x_i - x_c)}{r_y}, \\
    (|x|^{\alpha} + |y|^{\alpha})^{1/\alpha} - 1 &= 0
\end{align*}
$$

To warp from this ellipse to a superellipse, a point $(x_0, y_0)$ will be warped to $(x_f, y_f) = (x_0 + dx, y_0 + dy)$ with the following constraints and a new $\alpha > 2$:

$$
\begin{align*}
    x_f - x_c &= \lambda(x_0 - x_c), \\
    y_f - y_c &= \lambda(y_0 - y_c), \\
    (|x_0 - x_c|^2 + |y_0 - y_c|^2)^{1/2} - 1 &= 0, \\
    (|x_f - x_c|^\alpha + |y_f - y_c|^\alpha)^{1/\alpha} - 1 &= 0
\end{align*}
$$

Since we want this full warping at the edge of the feature, but don’t want it to affect the rest of the image, we scale $(dx, dy)$ depending on the points distance from the feature center. We use a continuous version of a scaled Poisson distribution with a peak value of 1 at a ellipse radius of 1.

The warp shift $(dx, dy)$ for each image pixel is found for the individual warp from all of the selected features in that image, and these shifts are all added for the final shift. The value of the pixels in the final image are then set to the value of the pixels in the original image at the corresponding shifted position.
Gradient Shading

Besides outputting the features of the image, the feature extraction algorithm also segments the image at the Voronoi diagram step. We can effectively simulate a brush stroke by changing the luminosity of the pixels in a particular region as a function of their distance to the stroke starting point, which is randomly selected as a point on the edge of the region. This doesn’t complete the brush stroke rendering effect. Appealing results are found with the function

\[ I_f = \frac{I_0}{1 + \left(\frac{R}{k}\right)^2} \]

where \( k \) is a decay length that is chosen to be twice the size of an average image region.

Non-salient Composition

To finish the image composition in the areas surrounding the chosen features, a nearest-neighbor approach is used: for each empty pixel, we find the closest pixel that has been filled with a feature. The value at the empty pixel is then set to the value of the pixel in the image from which the value at this neighboring feature pixel originates. In many tested cases, this nearest-neighbor approach works well, filling out features that have correctly been identified but whose size has been underestimated by the salience threshold step. Figure 3 shows an example output of this step.

Impressionist Rendering

Finally, with the final composition available, an impressionist rendering gives the image a true feeling of painted art using a method mainly adapted from Sparavigna and Marazzato [2]. An array of points, with randomly selected vertical and horizontal starting positions and point spacing, is found. For each point, a random vertical and horizontal shift, normally distributed
about zero shift, is selected. All of the pixels within some small radius of this shifted points have their values set to the value of the unshifted pixel. The method from this paper is slightly altered in that the standard deviation of the distribution from which the shift is selected and the radius chosen for the point stroke changes depending on the saliency of the region. For highly salient regions, details are kept by preferring small shifts and making small radii strokes. For regions of low saliency, large shifts and large strokes are acceptable. This process is repeated 20-100 times to fill in the picture with these “point” brush strokes. An added benefit of this procedure is an effective color quantization of the image, which was implemented in a different way by the previous Cubist rendering work.

Figure 4 displays several results of this final rendering step. Decent results are achieved with portrait image data sets. There does seem to be a higher tendency for extra eyes or extra mouths to appear, as compared to the previous work. For this method, portrait images aligned at the eye level tend to do better for eyes, while giving extra chins, and vice versa.
References


**Figure 1.** Left: original image. Center: Salience of the image, with feature points labeled. Right: Voronoi diagram segmenting the image, individual objects should appear near the center of these image segments and not get cut in half by the sections.
Figure 2. Left: result of the composition. Right: Colormap indicating the index of the image from which each feature originates. Features come somewhat evenly from all three source images, and create a composition that still keeps the overall shape of a face.
Figure 3. Left: Colormap indicating index of image from which each pixel will be sampled.

Right: final complete composition (with gradient directional stroke effects).
Figure 4. Example Cubist renditions.