# Dominant Color Learning by Subject Extraction

Kamelia Aryafar  Josh Attenberg  Fiona Condon

Etsy Inc., Data Science Group, Brooklyn, NY, USA

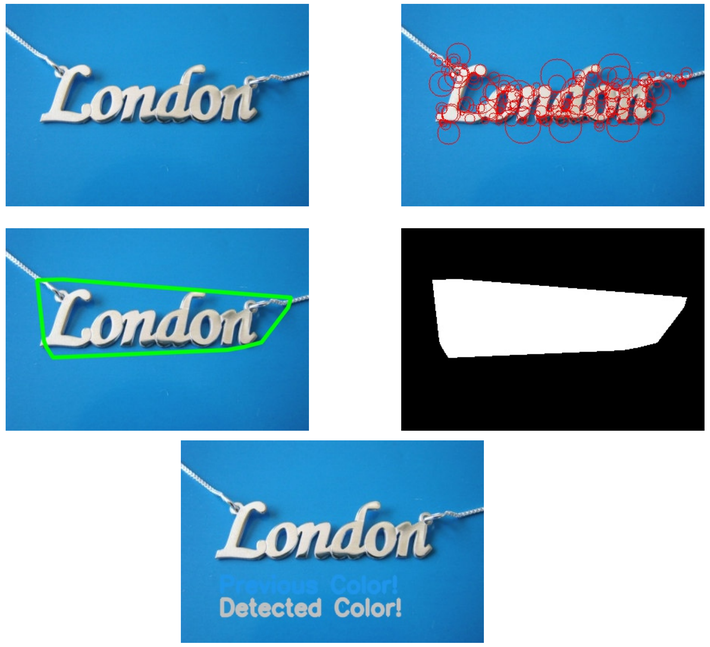
## Introduction

- This work proposes a method for dominant **color detection** in a large image data set from the Etsy website.
- Detecting the dominant color of an object in the image without any prior knowledge about the background model, the object characteristics or the scene geometry is a challenging problem.
- The main challenges in dominant color detection are:
  - The robust representation of image ROI (region of interest) that includes item of interest and isolation of the subject by background subtraction.
  - The extraction of dominant color from the approximated subject region.
- This work proposes the use of an automated method based on Canny's edge detection algorithm applying a threshold with hysteresis to identify the image ROI and extract the dominant color information.

## SURF Features Convex Hull Dominant Color Detection

We compare the proposed method to the convex hull of SURF features as a subject mask.Our experimental results show that detection of the contour of edges reduces the number of background pixels inside the obtained mask specifically when the objects are far apart within the region of interest.

- Set of SURF features $\{f_1, ..., f_\ell\}$ is used to present image descriptors.
- Convex hull of SURF features is used as the image mask to filter for ROI.
- Global maximum of RGB color histogram is used to assign dominant subject color.
- **Previous Color** denotes a dominant color detection result without subject detection while **Detected Color** denotes the result after subject detection.

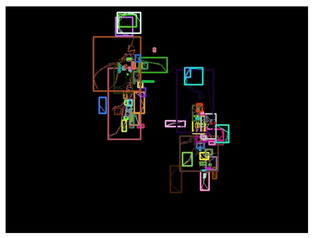


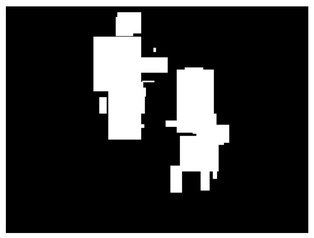

## Proposed Method-ROI Selection

- **ROI Selection and pre-processing:**
  - The raw image data is first convolved with a Gaussian filter to reduce noise sensitivity.
  - The subject extraction is then performed by first extracting the edges inside the image using Canny's edge detection algorithm applying a threshold with hysteresis.
  - Once the high and low thresholds are set properly, the image is represented as robust detected edges.

## Proposed Method-Dominant Color Detection

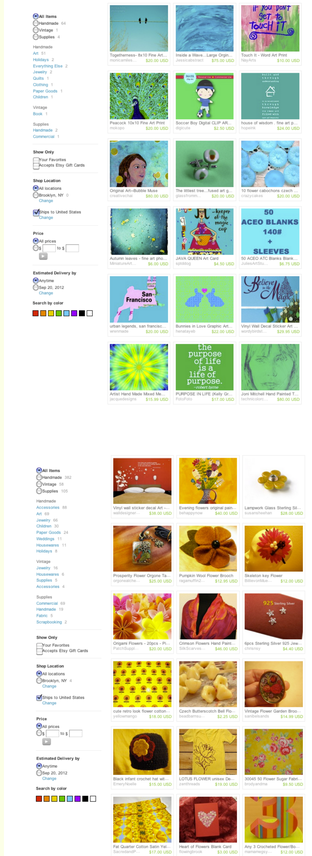
- **ROI Mask and Dominant Color Detection:**
  - We then create the contours of these edges to approximate an image mask accounting for objects that are far apart within the scene.
  - This approach minimizes the number of background pixels present inside the subject mask, thus a more accurate color representation is obtained.
  - A three dimensional color histogram is calculated for RGB channels of the masked image where the dominant color is presented as a strong maximum (spike).

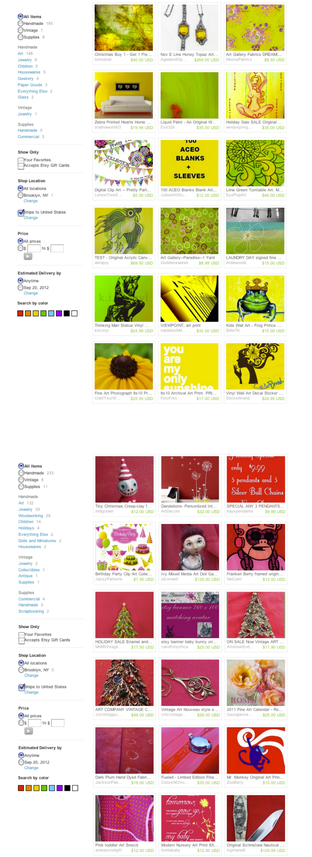





## Subject Color Filtering on Etsy

- **Experimental setup**
  - A filter by color is defined on Etsy website thresholding the RGB color values to 8 distinct color intervals (noise reduction)





## Future Work

- Classification of images based on subject color
  - Labeling subject colors using textual metadata (ground truth)
  - SVM classification on detected colors for images without metadata.
- Image similarity measures:
  - Using the proposed method as an ROI subject mask to refine image features (SIFT, SURF, PHOW)
  - Robust similarity detection
  - Minimize noise sensitivity
- Image duplicate detection:
  - Filter images with ROI mask to obtain image subject.
  - threshold feature vector similarity discarding background changes.

http://www.etsy.com/  http://www.cs.drexel.edu/  kca26@drexel.edu