Dataset for Identifying Tattoos

For our matching process, we want images that make sense given our specific use cases; we may try to match from images of just tattoos, we want our dataset to be of all cropped tattoos. Since we did not already have cropped tattoos from our true dataset, we used the NIST Tattoo Recognition Technology Challenge (Tatt-C) dataset \cite{grother}. These tattoos were captured during law enforcement agency operations, and thus the quality, exact content, environment, and process, may vary due to the nature of how they were caught. This is very similar to the data we want to deal with, in which we have no control over how the girls and their tattoos are captured. Thus, the Tatt-C dataset gives us a realistic representation of tattoos images we want to match.

Specifically, we focused on the tattoo identification section of the Tatt-C dataset, which includes 157 probes and 215 gallery images, with a total of 372 tattoo images. Each probe tattoo has anywhere from 1 to 10 possible matches in the gallery images. A match in this dataset is defined as two images of the same tattoo from the same subject collected during different encounters. Three examples are presented in Figure \ref{fig:tattoo\_identification}.

Tatt-c – what is it?

* Why we need cropped tattoos? – as is also specified in the Tatt-C paper, we want to have images that make sense given our specific use case.
* Same tattoo from same subject collected during different encounters
* “What is the retrieval performance for finding different instances of the same tattoo image from the same subject over time?”
* Still images of tattoos captured operationally by law enforcement agencies 🡪 many varying factors because of the nature of being captured operationally; not same environment, process, content, quality
* Five use cases: similarity, identification, region of interest, mixed media, tattoo detection
* 1 to 5 possible matches
* Person identification, finding people with the same tattoo
* Tattoo identification: 157 probes, 215 gallery images, total of 372 images
* Include details from/cite <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7126369&tag=1>
* Image with example

Matching Approaches

For a baseline approach, we decided to use a similarity metric based on the pixel values.

1. Nearest neighbor
   1. Baseline
   2. Resize black and white images to 16x16 square thumbnails
   3. Zero mean and normalize to adjust for brightness levels (and other things?)
   4. Euclidean distance between image pixel values to find most similar images
   5. Examples of what this worked well on/didn’t work on
   6. How pixels are bad representations (affected by rotation, scaling perspective, etc)
2. SIFT + Bag of Words
   1. Explain what it is, how SIFT is often used
   2. Motivation: identification but maybe also expand into similarity matching
   3. http://www.cse.msu.edu/biometrics/Publications/SoftBiometrics/LeeJinJainEnsembleRank\_ICPR10.pdf
3. SURF Nearest Neighbor

Testing Matching Robustness