Tattoo Recognition Technology - Challenge (Tatt-C): An Open Tattoo Database for Developing Tattoo Recognition Research

Mei Ngan and Patrick Grother
National Institute of Standards and Technology (NIST)
Gaithersburg, MD 20899
{mei,pgrother}@nist.gov

Abstract

Tattoos have been used for many years to assist law enforcement in investigations leading to the identification of criminals and victims. A tattoo is an elective biometric trait that could contain more discriminative information to support person identification than traditional soft biometrics such as age, gender and race. While some research has been done in the area of image-based tattoo detection and retrieval, it is not a mature domain. There are no common datasets to evaluate and develop operationally-relevant tattoo recognition applications. To address this shortcoming, the NIST Tattoo Recognition Technology Challenge (Tatt-C) database was developed as an initial tattoo research corpus that addresses use cases representative of operational scenarios.

The Tatt-C database represents an initial attempt to provide a set of ground-truthed tattoo images focused on, but not limited to, five primary use cases. This paper describes the details of the database along with the experimental protocols and test cases that should be followed, which will enable consistent performance comparison of tattoo recognition methods.

1. Introduction

This paper describes a database of tattoo images designed as an initial corpus to support and challenge the academic and commercial community in studying the problem of tattoo detection and retrieval. The Tatt-C dataset is free and instructions on obtaining the dataset can be found at the following web address: `http://www.nist.gov/itl/iad/ig/tatt-c.cfm`.

Recent surveys regarding the prevalence of tattooing indicates that one in five U.S. adults now has a tattoo [1]. While a tattoo alone may not be sufficient to differentiate between two individuals, it can help narrow the identity of suspects or victims in investigations. A tattoo is an elective biometric trait that could contain more discriminative characteristics and features for person identification than traditional soft biometrics such as age, gender, race, height, and weight [2]. As such, tattoos have been used for many years to assist law enforcement in the identification of criminals and victims and for investigative research purposes. Tattoos can provide valuable information on an individual's affiliations or expressed beliefs and can support identity verification of an individual. Historically, law enforcement agencies have followed the ANSI/NIST standard [3] to collect and assign keyword labels to tattoos. This keyword labeling approach comes with drawbacks, which include the limitation of ANSI/NIST standard class labels to describe the increasing variety of new tattoo designs, the need for multiple keywords to sufficiently describe some tattoos, and subjectivity in human annotation as the same tattoo can be labeled differently between examiners. For this reason, the shortcomings of keyword-based tattoo image retrieval have driven the need for automated image-based tattoo detection and retrieval capabilities.

While some research has been done in the area of image-based tattoo recognition [4, 5, 6, 7, 8, 9], it is not a mature industry. There is no common research data and use cases to evaluate and develop systems for next generation government applications. To address this shortcoming, the Tatt-C dataset was developed as an initial tattoo research corpus that addresses use cases that represent operational scenarios.

The Tatt-C dataset consists of still images of tattoos captured operationally by law enforcement agencies. The operational nature of this corpus imposes challenges on traditional image retrieval methodologies given the large variation in capture environment/process and tattoo content/quality. The following are examples of such challenges represented in the Tatt-C dataset:

- Varying image lighting and scale
- Partial occlusions from clothing and noisy image backgrounds

- Different tattoo background or embellishments around primary tattoo content
- Blended images or multiple tattoos in a single image
- Inconsistent orientation of body/appendages and images
- Extremely faded tattoos
- Ambiguous or unfamiliar abstractions (difficult or impossible for different people to interpret and describe consistently)

2. Related Work

2.1. Existing Databases

At the time this paper was written and to the best of our knowledge, there currently does not exist a public tattoo database with ground truth data on a relatively large number of images that is available from the Tatt-C corpus. Much of the tattoo recognition work published in the academic literature tested their methods on images collected from the Internet, but are not publicly available in consolidated form and/or operational images from local law enforcement, which only the authors had access to.

Prior publications [10], [4], and [11] tested with derivatives of 100, 2157, and 4323 tattoo images, respectively, originally collected from the Internet. [6] also published results on a dataset of 256 tattoo images collected from the Internet, and [8] evaluated on a dataset of 18,922 images collected from the Internet and sampling from public face databases. [11] tested against 69,507 operational tattoo images extracted from the Michigan State Police Tattoo Database, which is not available for public use. [12] had access to 63,593 operational tattoo images from an undisclosed source for testing. [7] had access to a set of 327,049 tattoo images collected by the German police. [5] tested with 4,000 prison/gang tattoo images, derived from 100 original tattoo images from an undisclosed source. [9] tested their algorithm on a dataset of 444 tattoo images gathered from the Royal Thai Police's Criminal Records Division. [13] constructed a small database of 100 tattoos and their corresponding sketches drawn by two subjects. While this sketch dataset is available for public download, it is primarily focused on matching sketches to tattoos.

2.2. Standardized Challenges

There have been a number of datasets with "a take-home test" protocol similar to Tatt-C: The NIST Multiple Biometric Grand Challenge (MBGC) [14] for face and iris recognition, Labeled Faces in the Wild (LFW) [15] for unconstrained face recognition, and NIST i-vector Machine Learning Challenge [16] for machine learning for use in speaker recognition. These datasets along with Tatt-C help promote development, advance the state-of-the-art, and benchmark progress in computer vision applied to different areas of biometrics.

3. Use Cases

The Tatt-C dataset provides a basis for objectively measuring and comparing tattoo recognition capabilities, with partitions focused on but not limited to the use cases detailed in this section. Table 1 provides a summary of the use cases.

3.1. Tattoo Similarity

Challenge: Given an image of a tattoo, find visually similar or related tattoos from a database. This use case has application in group or gang affiliation, for example, in cases of identifying individuals in a criminal gang that could lead to other individuals with very similar tattoos that are very likely in the same gang. The test data for this use case is composed of groups of visually similar tattoos from different subjects collected at different times. For any given probe image, there are one or more correctly matching gallery image(s) in the database. Examples of sets of visually similar tattoos are shown in Figure 1.

Figure 1. Examples of image sets from the Tattoo Similarity use case.

3.2. Tattoo Identification

Challenge: Given an image of a tattoo, find different instances of the same tattoo from the same subject from a database. This use case has application in investigation supporting identification of an individual, for example, in the case of a robbery where the suspect is wearing a mask and gloves and the surveillance camera may be able to record a tattoo exposed on the neck or the arm from the suspect. The test data for this use case is composed of images of the same tattoo from the same subject collected during different encounters. For each probe image, there could be one or more correctly matching tattoo image(s) in the database. Examples of same-subject tattoos are shown in Figure 2.

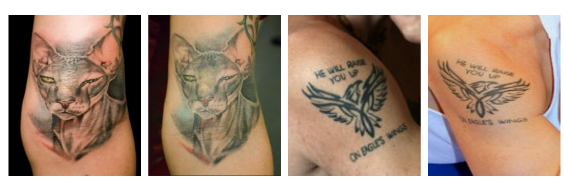

Figure 2. Examples of image sets from the Tattoo Identification use case.

	Tattoo Similarity	Tattoo Identification	Region of Interest	Mixed Media	Tattoo Detection
Use Case	Match visually similar or related tattoos from different subjects	Match different instances of the same tattoo from the same subject over time	Match small region of interest contained in a larger tattoo	Match visually similar or related tattoos across different mediums	Detect whether an image contains a tattoo
Utility Example	Group Affiliation	Person identification	Person identification	Intelligence gathering	Database construction and maintenance
Task	One-to-many search	One-to-many search	One-to-many search	One-to-many search	Classification
Types of images	Tattoos	Tattoos	Tattoos	Tattoos, sketches, graffiti, computer graphics	Tattoos, faces
Number of images	2212	372	454	454	2349
Number of probes	851	157	297	181	2349

Table 1. Summary of Tatt-C use cases.

3.3. Region of Interest

Challenge: Given a small region from an image of a tattoo, find the larger tattoo image from which the smaller region is contained from a database. This use case has application in investigation supporting identification of an individual, for example, in the case where a suspect's face is occluded but the surveillance camera is able to capture a portion of a tattoo that is exposed on the arm. The test data for this use case is composed of multiple tattoo sections (probes) that belong to a larger tattoo canvas (gallery image) from the same subject collected during different encounters. For each probe image, there exists one correctly matching tattoo image in the database. Region of interest examples are shown in Figure 3.

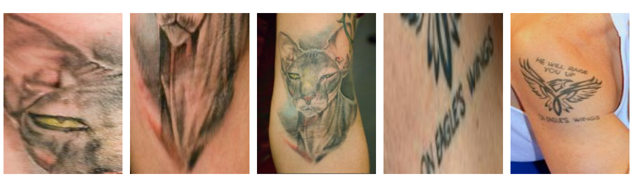

Figure 3. Examples of image sets from the Region of Interest use case.

3.4. Mixed Media

Challenge: Given other (non-tattoo) types of images, (e.g. sketches, scanned print, computer graphics, graffiti, etc.), find visually similar or related tattoos from a database. This use case has application in investigative intelligence gathering. In cases where the image of the tattoo is not captured on a camera, but an individual witnessed someone with a tattoo involved in criminal activity, a description of the tattoo can be provided to a forensic artist, and the sketch can be searched against a database for potential matches. The test data for this use case is composed of groups of visually similar mixed media images and tattoos. For any given mixed media probe image, there are one or more correctly matching tattoo image(s) in the database. Mixed media examples are shown in Figure 4.

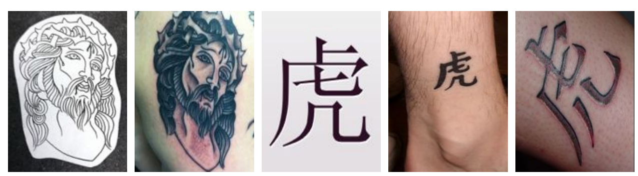

Figure 4. Examples of image sets from the Mixed Media use case.

3.5. Tattoo Detection

Challenge: Given an image, detect whether it contains a tattoo or not. This use case has application in database construction and maintenance when large amounts of unlabeled data is comingled, making automatic extraction of different types of images a challenge. An example is the ANSI/NIST Type 10 record where facial mugshot images and scar, mark, tattoo (SMT) images are stored in the same record, and with a percentage of the data mislabeled or not labeled at all, it presents a challenge to automated extraction of the data based on image content. The test data for this use case is composed of tattoo images and face images extracted from the Multiple Encounter Database 2 (MEDS-II) [17], as shown in Figure 5.

#	ID	Use Case	Probes	Gallery	Enroll Background	Use Metadata
1	SIM-1	Tattoo Similarity	original	original	no	no
2	SIM-2	Tattoo Similarity	original	original	yes (original)	no
3	SIM-1-CR	Tattoo Similarity	cropped	cropped	no	no
4	SIM-2-CR	Tattoo Similarity	cropped	cropped	yes (cropped)	no
5	SIM-1-CR-PROBES	Tattoo Similarity	cropped	original	no	no
6	SIM-2-CR-PROBES	Tattoo Similarity	cropped	original	yes (original)	no
7	ID-1	Tattoo Identification	original	original	no	no
8	ID-2	Tattoo Identification	original	original	yes (cropped)	no
9	ROI-1	Region of Interest	original	original	no	no
10	ROI-2	Region of Interest	original	original	yes (cropped)	no
11	MM-1	Mixed Media	original	original	no	no
12	MM-2	Mixed Media	original	original	yes (original)	no
13	DET-1	Tattoo Detection	original	NA	NA	no
{1-12}	{*}-META	*	*	*	*	yes

Table 2. Summary of Tatt-C experimental design.

Figure 5. Examples of images from the Tattoo Detection use case.

4. Test Protocol

Table 2 shows the list of experiments for the Tatt-C dataset. For the one-to-many search tasks (i.e., 1-12), algorithms would enroll the gallery images plus background images (if applicable), then search with the probe images. For the classification task (i.e., 13), algorithms would, for each probe, classify whether the image contains a tattoo or not.

4.1. Cross-Validation

Within each use case, the data has been partitioned into five mutually exclusive subsets. Performance results should be produced using a 5-fold cross-validation scheme, that is, within each use case, participants should conduct five separate experiments in a rotational cross-validation scheme. In each experiment, four of the subsets should be combined to form a training set, with the fifth subset used for testing. For example, the first experiment would use subsets (2, 3, 4, 5) for training and subset 1 for testing. The fourth experiment would use subsets (1, 2, 3, 5) for training and subset 4 for testing. Algorithm parameters under each experiment should be set using only the training data for that experiment to avoid fitting to the test data. In other words, each of the five experiments (both the training and test phases) should be run completely independently of the others.

4.2. Outside Training Data

The use of data outside of the Tatt-C dataset for training purposes is allowed. Participants will be asked to disclose the use of external training data upon results submission to Tatt-C, hence submitting to one or the other of the following tracks:

1) Provided data
2) Outside data

with 1) for submissions using only Tatt-C images and labels and 2) for submissions using any outside images or labels for training.

4.3. Background Images

In addition to the probes and their corresponding mates in the gallery, there are 4,332 background tattoo images used for padding the enrollment gallery for the various one-to-many search test cases where specified.

4.4. Metadata Utilization

Within each use case folder is a `metadata.txt` file that contains the following image information:

- `img_name`: name of the image
- `ansi_nist_class`: the general class and subclass code chosen from the eight class codes and seventy subclass codes specified in the ANSI/NIST standard
- `description`: free-text field that provides additional qualifiers to describe the image
- `color`: the color(s) of the tattoo as specified by the ANSI/NIST standard
- `body_location`: general location of the tattoo as specified by the ANSI/NIST standard, referencing the

U.S. National Crime Information Center (NCIC) SMT Body Location Codes
- `rect_coordinates(x,y,width,height)`: coordinates for bounding box drawn around tattoo content
- `orientation`: orientation specification of the tattoo image

While Tatt-C is primarily focused on image-based tattoo retrieval, it is of interest to compare the performance impact of leveraging metadata (i.e., labels) where available. The last line item in Table 2 allows experiments 1-12 to be repeated with the additional use of metadata where available.

4.5. Original vs. Cropped Images

Many of the original tattoo images collected operationally contain clothing and background distractions, which poses challenges to traditional image retrieval methodologies. Cropped versions of the original images where most of the background distractions are removed, leaving the primary tattoo content in the image, are available for the Tattoo Similarity use case and the background images. There are some experiments from Table 2 that specify the use of cropped versions of these images, and the goal is to measure the performance impact of removing the background distractions in the tattoo images. Where background images are used to pad the enrollment gallery, some experiments specify the use of the original tattoo images with background distractions while others use the cropped images. This is primarily driven by the image content for the particular use case, meaning if most of the gallery images contain images that contain distractions, then a background with similar images properties would be desired for consistency and to prevent filtering of the gallery based on clothing or background rather than tattoo content.

5. Dataset Construction

5.1. Tattoo Similarity

The data for the Tattoo Similarity use case was prepared following a two-phased ground-truth methodology.

Phase 1: Establishing Semantic Ground Truth
A total of 200,000 operational tattoo images were provided to a group of ten biometric examiners. Examiners assigned labels to characterize the content of the tattoo in terms of ANSI/NIST standard classes/subclasses and customized labels. The labeling phase also included capture of the tattoo's orientation, color, boxed region around primary tattoo content, and body location. Images could contain more than one tattoo. In those cases the examiner boxed the tattoos separately. Images could also contain tattoos requiring more than one label, however for the scope of this corpus images are only grouped within one label.

When the ANSI/NIST labels fail to accurately describe an image, examiners can enter custom labels. A total of twenty-six pre-populated customized labels were made available to the examiners based on some familiarity with the data. Examiners were encouraged to use these labels whenever possible for consistency but could also create their own at any time, especially if the pre-populated labels were found unrepresentative of the data.

The labeling phase was conducted over a four-week period with ten examiners working on a part-time basis. A total of 15,798 labeled images were generated in that timeframe. For the Tatt-C corpus, images where examiners boxed more than one tattoo were excluded from the visual ground-truthing phase.

Phase 2: Establishing Visual Ground Truth
Of those labeled, approximately 50% or 7,932 grouped among 142 semantic labels were used for establishing visual ground-truth. Visual similarity of images can be very subjective. Examiners were asked to weigh, on a scale of 1 (low similarity) to 5 (high similarity) how visually similar they perceived a pair of images to be based on content, composition, style, and details to make the determination as objective as possible. Examiners were asked to review all possible combination of images that were generated from within the same semantic label group. Examiners were provided 819,631 thousand unique pairings to evaluate, based on the unique combinations from the 142 labels. Each examiner's workload was randomized, so some pair scores represent more than one examiner's opinion. Ten biometric examiners scored 43% (352,441) of the total pairings over a four-week period. Of the image pairings assessed for visual similarity, nearly 45% were scored by more than one examiner. Despite sharing the same semantic label, very few of the image pairs out of the total scored pairings were visually similar, reinforcing the large intra-class variation problem. Image pairs that had a visual similarity score of 3 or above and an average score difference between multiple examiners of 1 or less (where applicable) were incorporated into the Tatt-C dataset, resulting in 851 unique probes with one or more matching gallery image(s).

5.2. Tattoo Identification

The data for this use case was prepared by two examiners using a manual process of finding subjects with multiple encounters where pictures of the same tattoo were collected. Given many of these images contained personally identifiable information (PII) (e.g. faces, arm-bands with names, etc.), the majority of the images were cropped to eliminate any PII, resulting in 157 unique probes with one or more matching gallery image(s).

5.3. Region of Interest

The data for this use case was prepared by one examiner using a manual process of taking each probe image from the Tattoo Identification use case and cropping out specific regions to form a new probe set of unique tattoo sections. One gallery image that corresponded to the original probe was selected to be the matching image to be retrieved from the database. This resulted in 297 unique probes with one and only one matching gallery image.

5.4. Mixed Media

The data for this use case consists of sketches, computer graphics, and graffiti images that correspond to known tattoo image(s). There are 100 tattoo sketches drawn by an examiner, each corresponding to a known tattoo image. The protocol for drawing the tattoo sketch was similar to the process exercised in [10]. The tattoo image was presented to the examiner for as long as they needed. Thirty seconds later, the examiner was asked to draw a tattoo sketch on white paper from memory. The remainder of the data originated from closed case files supplied by law enforcement units, resulting in 181 unique mixed media probes with one ore more matching gallery tattoo image(s).

5.5. Tattoo Detection

For the data in this use case, 1,349 images were selected from a set of 15,798 known tattoos. Approximately 50% of the selected tattoo images contain some form of a face (e.g. clown, skull, Jesus head, etc.). The remainder of the tattoos were randomly selected. The dataset is comingled with 1,000 face images extracted from the MEDS-II dataset, resulting in 2,349 images for this use case.

6. Summary

Tatt-C is the first publicly available tattoo database focused on five primary operational use cases, with goals to

1) provide a seminal research corpus of operationally-relevant tattoos to evaluate and advance tattoo recognition methods,
2) allow for comparison of tattoo recognition methods via a common database and protocol, and
3) assess the current state-of-the-art in tattoo recognition.

References

[1] S. Braverman. One in Five U.S. Adults Now Has a Tattoo. *The Harris Poll*, 2012. http://www.harrisinteractive.com/NewsRoom/HarrisPolls/tabid/447/mid/1508/articleId/970/ctl/ReadCustom%20Default/Default.aspx.

[2] J.-P. Beauthier, P. Lefevre, and E. D. Valck. Autopsy and Identification Techniques, The Tsunami Threat - Research and Technology. In *N.-A. Mörner, ed., InTech*, 2011.

[3] ANSI/NIST-ITL 1-2011:Update 2013, NIST Special Publication 500-290, Data Format for the Interchange of Fingerprint, Facial and Other Biometric Information. http://www.nist.gov/itl/iad/ig/ansi_standard.cfm.

[4] A. K. Jain, J.-E. Lee, and R. Jin. Tattoo-id: automatic tattoo image retrieval for suspect and victim identification. In *PCM '07*, 2007.

[5] S. T. Acton and A. Rossi. Matching and retrieval of tattoo images: Active contour cbir and glocal image features. In *Proc. of the 2008 IEEE Southwest Symposium on Image Analysis and Interpretation*, pages 21–24, 2008.

[6] J. D. Allen, N. Zhao, J. Yuan, and X. Liu. Unsupervised tattoo segmentation combining bottom-up and top-down cues. In *Proc. of the SPIE 8063, Mobile Multimedia/Image Processing, Security, and Applications 2011*, 2011.

[7] D. Manger. Large-scale tattoo image retrieval. In *Proc. of the Conference on Computer and Robot Vision*, pages 454–459, 2012.

[8] B. Heflin, W. J. Scheirer, and T. E. Boult. Detecting and classifying scars, marks, and tattoos found in the wild. In *BTAS '12*, pages 31–38, 2012.

[9] P. Duangphasuk and W. Kurutach. Tattoo skin detection and segmentation using image negative method. In *Proc. of the 13th International Symposium on Communications and Information Technologies*, pages 354–359, 2013.

[10] A. K. Jain, Y. Chen, and U. Park. Scars, marks and tattoos: Physical attributes for person identification. In *MSU Tech Report CSE 07-22*, 2007.

[11] J.-E. Lee, A. K. Jain, and R. Jin. Scars, marks and tattoos (SMT): Soft biometric for suspect and victim identification. In *Proc. of the Biometrics Symposium*, pages 1–8, 2008.

[12] A. K. Jain, J.-E. Lee, and N. Gregg. Content-based image retrieval: An application to tattoo images. In *Proc. of the International Conference on Image Processing*, pages 2745–2748, 2009.

[13] H. Han and A. K. Jain. Tattoo based identification: Sketch to image matching. In *Proc. of the International Conference on Biometrics Compendium*, pages 1–8, 2013.

[14] P. J. Phillips et al. Overview of the multiple biometrics grand challenge. In *Advances in Biometrics*, volume 5558 of *Lecture Notes in Computer Science*, pages 705–714. Springer Berlin Heidelberg, 2009.

[15] G. B. Huang, M. Ramesh, T. B., and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.

[16] C. S. Greenberg et al. The NIST 2014 Speaker Recognition i-Vector Machine Learning Challenge. In *Odyssey 2014: The Speaker and Language Recognition Workshop*, pages 224–230, 2014.

[17] NIST Special Database 32 - Multiple Encounter Dataset 2 (MEDS-II), NISTIR 7807. http://www.nist.gov/itl/iad/ig/sd32.cfm.