**The author(s) shown below used Federal funding provided by the U.S. Department of Justice to prepare the following resource:**

| | |
|---|---|
| **Document Title:** | **Automation-Supported Curation of Large Forensic Image Databases** |
| **Author(s):** | **Audris Mockus, Ph.D., Dawnie Wolfe Steadman, Ph.D., D-ABFA** |
| **Document Number:** | **306559** |
| **Date Received:** | **May 2023** |
| **Award Number:** | **2018-DU-BX-0181** |

Department of Justice, Office of Justice Programs
National Institute of Justice
Grant # 2018-DU-BX-0181

# Automation-Supported Curation of Large Forensic Image Databases

Submitted by:

Audris Mockus, Ph.D.
Ericsson-Harlan D. Mills Chair Professor of
Digital Archeology and Evidence Engineering
865-974-2265; audris@utk.edu

Dawnie Wolfe Steadman, Ph.D., D-ABFA
Director of the Forensic Anthropology Center
Professor of Anthropology
865-974-0909; osteo@utk.edu

DUNS: 00-388-7891
EIN: 162-6001636A1

The University of Tennessee
1 Circle Park Drive
Knoxville, TN 37996-0003

Recipient Account: #R011344646

Final Report

# Purpose and Objectives of the Project

The Forensic Anthropology Center (FAC) at the University of Tennessee, Knoxville (UTK), is home to the Anthropology Research Facility (ARF), where human decomposition has been studied since 1981. Since 2012, multiple photographs of each donor (approximately 100 donors annually) have been taken daily throughout the entire decomposition process. To date, the database occupies more than 4TB of disk space and contains over one million photographs obtained from more than 500 donors.

It is exceedingly difficult to use such an enormous collection for research or law enforcement due to its sheer size and inability to access the forensically relevant features within the image content, suggesting that emerging big data approaches might help. Within biological, and subsequently, forensic, anthropology, publicly available datasets that would be considered as big data are lacking. Particularly, we are only aware of a single freely available photographic database that includes multiple, daily, photographs of hundreds of cadavers that chronicle the decomposition process from fresh to skeletal. This database was created because of our initial research (NIJ Grant # 2016-DN-BX) which produced the ICPUTRD: Image Cloud Platform for Use in Tagging and Research on Decomposition. ICPUTRD allows forensic researchers to browse, search, and annotate this enormous collection. We refer to the process of image selection and tagging for a particular research study or for a particular use as curation. By tagging we mean identifying a specific subarea of the image that contains a feature related to a specific term from the nomenclature. For example, an area with scavenging, larvae, discoloration, or purge. The ICPUTRD platform has begun to organize, standardize and create a valuable research tool in forensic science, several obstacles had to be removed to realize its full potential. Specifically, it developed an initial nomenclature with standard human decomposition-relevant concepts. However, the curation effort was exceedingly time-consuming and the work under the initial grant was able to partially tag only approximately 1000 photos or less than one tenth of one percent.

The work supported by this grant aimed to develop methods that would at least partially automate the curation. Specifically, we proposed to a) train deep learning models on the set of existing tags and use these trained models to tag remaining one million images; b) implement new capabilities in ICPUTRD that make it possible for experts in human decomposition to evaluate and improve the accuracy of these model-generated tags with minimal effort; c) organize the detailed multivariate-temporal data representing the incidence of the features representing the nomenclature terms and related covariates (such as temperature/humidity exposure) for hundreds of donors and provide relevant analysis tools and methods, thus exposing detailed quantitative forensic data encoded in millions of images.

## Project Design

The proposed research was divided into three components:

1) develop a supervised learning approach that is fitted using over a thousand tagged images and that predicts standard nomenclature tags for the remaining one million untagged images.

2) implement capabilities to manually validate the accuracy of automatic tagging including developing active learning approaches to minimize the amount of such manual effort.

3) develop an analytic database of spatiotemporal tag counts for the whole collection of images to quantify the human decomposition process on an unprecedented scale.

## Results

In short, we have achieved the primary goals set out in the proposal.

1. Developing supervised learning approach

This required doing several steps. Increasing the accuracy of the tags using polygonal masks, increasing the sample of images with rare types of nomenclature, develop models that group images of the same body part taken at separate times, and developing segmentation models for the body parts.

As the first step we needed to refine thousands of rectangular masks used to select areas of images representing nomenclature concepts using more precise polygonal masks. We have implemented necessary software improvements and refined over 5000 of previously tagged rectangular masks.

Second, we fitted the models for rare nomenclature traits to predict images that are more likely to have them (some traits occurred in less than one in one thousand images). This step resulted in a significant increase in the number of tagged images, but we also observed significant confounding from the background that often has textures and colors closely resembling some of the important nomenclature concepts. We therefore refocused our subsequent efforts on creating deep learning models that segment the body parts. To accomplish that, we first needed to identify body parts automatically without needing an exceedingly large number of manually tagged images.

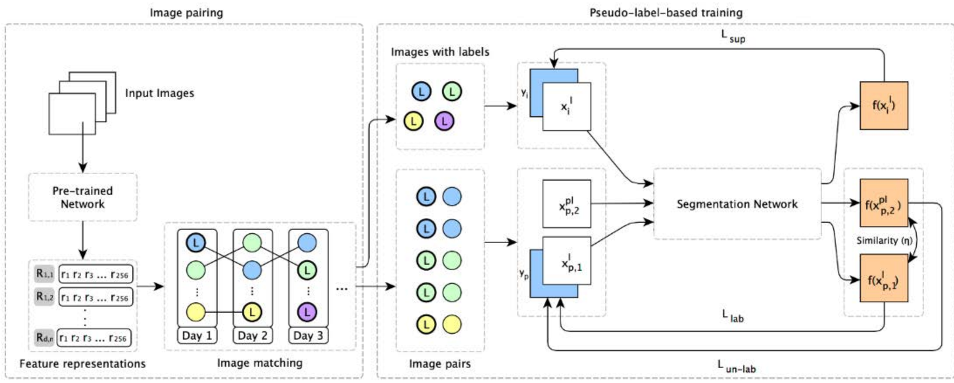
The following algorithms we have developed are described in more detail below.

**Semantic Clustering via Image Sequence Merging (SChiSM**) aims to make the ICIPUTRD collection more accessible to forensic experts by identifying various body parts and tracing them through their evolution despite their distinct appearances at various stages of decay from "fresh" to "skeletonized". We developed an unsupervised technique for clustering images that builds sequences of similar images representing the evolution of each body part through stages of decomposition. Evaluation of our method on 34,476 human decomposition images shows that our method significantly outperforms the state-of-the-art clustering method in this application.

**Pseudo Pixel-level Labeling** for images with evolving content also leverage the evolving nature of images depicting the decay process in human decomposition data to design a simple yet effective pseudo-pixel-level label generation technique to reduce the amount of effort for manual annotation of such images. We first identify sequences of images with a minimum variation that are most suitable to share the same or similar annotation using an unsupervised approach. Given one user-annotated image in each sequence, we propagate the annotation to the remaining images in the sequence by merging it with annotations produced by a state-of-the-art CAM-based pseudo label generation technique. To evaluate the quality of our pseudo-pixel-level labels, we train two semantic segmentation models with VGG and ResNet backbones on images labeled using our pseudo labeling method and those of a state-of-the-art method. The results indicate that using our pseudo-labels instead of those generated using the state-of-the-art method in the training process improves the mean-IoU and the frequency-weighted-IoU of the VGG and ResNet-based semantic segmentation models by 3.36%, 2.58%, 10.39%, and 12.91% respectively.

**Similarity-based Label Reuse for semantic segmentation (SLRNeT**) is a simple yet effective semantic segmentation technique to reduce the amount of effort for manual annotation of such images. In SLRNet, we first identify best matches for the annotated images from the pool of unannotated images using an unsupervised algorithm. Images with a minimum variation that are most suitable to share the same or similar annotation to the annotated images are considered as good matches. We then generate pairs of annotated and unannotated images from the found matches such that each annotated image is paired with all unannotated images in its matched pool. Some example pairs are shown in the following figure.

Once the pairs are created, we reuse the annotation of the annotated image as a pseudo-label for the unannotated image in each pair. These pairs along with the original annotated images are then used to train a semantic segmentation network with a custom loss function designed to control the impact of the pseudo-labels in the training process. The overall architecture of SLRNet is shown in the following figure.




We evaluate our method on the human decomposition images and compare our method with two state-of-the-art semi-supervised semantic segmentation methods: CCT (Ouali  et.al. ) and PseudoSeg (Zou et. al. ). Results indicate that SLRNet, while having a much simpler conceptual structure and correspondingly shorter runtime, outperforms both CCT and PseudoSeg on images of human decomposition by 9.95% and 10.23% in terms of mean IoU (intersection over Union of the segmentations), and 3.37 and 3.05 in terms of mean Accuracy.

2. Capabilities to manually validate the accuracy of automatic tagging

As described in the previous goal, we needed a way to validate the unsupervised image clustering. To do that, we expanded the functionality of ICPUTRD to enable manual feedback on this automatic grouping. The workflow starts from grouping a batch of images according to a particular version of the clustering algorithm and then presenting that batch for manual inspection by an expert. The task of the expert is to mark images that do not belong to a cluster. This is a quite different task than annotating a single image and, to make the expert productive, requires presenting that expert with many images and allowing them to click on those that do not belong to the cluster, which increases efficiency.  We integrated and evaluated this newly developed clustering approach (PLUD) with ICPUTRD. As a result of using it for classification and sequencing, and our labeling, we now have 55489 image-level labels. PLUD helps to employ contemporary supervised image analysis methods on image data, by first cleaning and organizing, and then manually labeling for the nomenclature employed in the specific domain, which is a time consuming and expensive endeavor. PLUD provides an iterative semi-supervised workflow to minimize the effort spent by an expert and handles realistic large collections of images. PLUD is an iterative sequence of unsupervised clustering, human assistance, and supervised classification. With each iteration 1) the labeled dataset grows, 2) the generality of the classification method and its accuracy increases, and 3) manual effort is reduced. We evaluated the effectiveness of our system, by applying it on over a million images documenting human decomposition. In our experiment comparing manual labeling with labeling conducted with the support of PLUD, we found that it reduces the time needed to label data and produces highly accurate models for this domain.

3. Analytic database of spatiotemporal tags

To make the ICPUTRD database more accessible, we need to ensure the privacy of the donors. One of the important tasks is to exclude any references to the donor ID and year when the donor was photographed. To do that we have developed an indexing scheme that removes information about the original donor ID and date and replaces it with a random ID that includes season. This required several changes to the naming and organization of the images and required a migration of the database of tags. With the new scheme in place, we expect to deploy the latest version of the software during this period.

We also started experiments where we created a database that linked weather summaries as experienced from deposition to the time the photograph was taken to be able to calculate forensically relevant quantities typically associated with time of death estimation.

## Implications for Criminal Justice Policy and Practice in the U.S.

Estimating decomposition events is central to forensic science and the criminal justice system in the United States. Previous research (Megyesi et al., 2005) has taken the requisite steps towards quantifying the decomposition process, though unaccounted for variability remains. Applying a big data approach for more than one million images collected in ARF is likely to dramatically increase the fidelity to the analysis and yield more accurate results. The results of our study have potential to directly impact the medicolegal community by providing opportunities to model the decomposition process, accounting more precisely for more sources of error (scavenging events, for example), and produce known error rates. The legal value of the evidence is increased when there are associated, quantifiable error rates (Christensen and Crowder 2009).

The proposed research created deep learning models needed to accurately group images by body parts, to segment body parts, and to achieve accurate results with few manually tagged images. These models provide a scientific basis to segment images not just in ARF, but in any other similar facility, or, potentially, for photos taken during casework.

## Scholarly Products

Conference Papers Presented:

- Sara Mousavi, Audris Mockus, Dawnie W. Steadman, Angela M. Dautartas. *Machine Learning to Detect and Localize Forensics-Relevant Features*. Paper presented at the 71st annual scientific meeting of the American Academy of Forensic Sciences, Baltimore, Maryland, February 22, 2019.
- Audris Mockus, *Image Cloud Platform for Use in Tagging and Research on Decomposition* at the NIJ R&D symposium co-located with AAFS 2019.

Publications (Accepted and in Preparation):

- Mousavi, Sara and Nabati, Ramin and Kleeschulte, Megan, Steadman, Dawnie, and Mockus, Audris, *Machine-assisted annotation of forensic imagery*, 26th IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, September 22-25, 2019.
- Mousavi, S., Lee, D., Griffin, T., Steadman, D., & Mockus, A. (2020, October). Collaborative learning of semi-supervised clustering and classification for labeling uncurated data. In 2020 IEEE International Conference on Image Processing (ICIP) (pp. 1716-1720). IEEE.
- Mousavi, S., Lee, D., Griffin, T., Cross, K., Steadman, D., & Mockus, A. (2021). SChISM: Semantic Clustering via Image Sequence Merging for Images of Human-Decomposition. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 2190-2199).
- Mousavi, S, Zhenning yang, Kelley Cross, Dawnie Steadman, Audris Mockus, “Similarity-based Label Reuse for Semi-Supervised Semantic Segmentation of Human Decomposition Images”, Submitted to IEEE Winter Conference on Applications of Computer Vision *(WACV 2022)*.
- Mousavi, S., Lee, D., Griffin, T., Steadman, D., & Mockus, A. (2020, October). Collaborative learning of semi-supervised clustering and classification for labeling uncurated data. In 2020 IEEE International Conference on Image Processing (ICIP) (pp. 1716-1720). IEEE.
-

Other publications

- Mousavi, S., Lee, D., Griffin, T., Steadman, D., & Mockus, A. (2019). An analytical workflow for clustering forensic images. arXiv preprint arXiv:2001.05845.
- Mousavi, S., Yang, Z., Cross, K., Steadman, D., & Mockus, A. (2021). Pseudo Pixel-level Labeling for Images with Evolving Content. arXiv preprint arXiv:2105.09975.
- Mousavi, S., Lee, D., Griffin, T., Steadman, D., & Mockus, A. (2020, April). An Analytical Workflow for Clustering Forensic Images (Student Abstract). In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 10, pp. 13879-13880).
- Sara Mousavi. *Auto-curation of Large Evolving Image Datasets*. PhD thesis, The University of Tennessee, Knoxville, TN, 2021.

Software and documentation and other presentations:

- ICPUTRD platform source code and documentation: https:/bitbucket.org/icputrd/platform
- ICPUTRD platform tag development: https:/bitbucket.org/icputrd/tags

**References:**

Christensen AM, Crowder CM. Evidentiary standards for forensic anthropology. Journal of forensic sciences. 2009 Nov;54(6):1211-6.

Megyesi MS, Nawrocki SP, Haskell NH. 2005. Using accumulated degree days to estimate the postmortem interval from decomposed human remains. Journal of Forensic Sciences 50(3):618–26.

Yassine Ouali, C´eline Hudelot, and Myriam Tami. Semi-supervised semantic segmentation with cross-consistencytraining. InProceedings of the IEEE/CVF Conference onComputer Vision and Pattern Recognition, pages 12674–12684, 2020

Yuliang Zou, Zizhao Zhang, Han Zhang, Chun-Liang Li,Xiao Bian, Jia-Bin Huang, and Tomas Pfister. Pseudoseg:Designing pseudo labels for semantic segmentation.arXivpreprint arXiv:2010.09713, 2020