STAGE OF DECAY ESTIMATION EXPLOITING EXOGENOUS AND ENDOGENOUS IMAGE ATTRIBUTES TO MINIMIZE MANUAL LABELING EFFORTS AND MAXIMIZE CLASSIFICATION PERFORMANCE

Anna-Maria Nau[1,2] *Audris Mockus*[1] *Dawnie Wolfe Steadman*[3]

[1] Department of Electrical Engineering & Computer Science, University of Tennessee, Knoxville, TN, USA
[2]The Bredesen Center, University of Tennessee, Knoxville, TN, USA
[3]Department of Anthropology, University of Tennessee, Knoxville, TN, USA

ABSTRACT

Establishing the stage of decay for human remains is an important and common task in forensic anthropology, which is presently done manually by trained experts. We aim to develop a supervised vision system, leveraging a large human decomposition image dataset, to perform stage of decay estimation. To minimize the manual labeling efforts and costs, we introduce and evaluate a label propagation method that exploits both exogenous and endogenous image attributes in conjunction with domain knowledge. The proposed label propagation method performed up to twice as fast (100%) as the baseline method and improved the classification macro-averaged F1 score of the best model from 0.763 to 0.863 (13.1%) when using three stage of decay classes and from 0.695 to 0.803 (15.54%) when using four stage of decay classes.

***Index Terms*—** Image classification, Transfer learning, Human decay modeling, Forensic science

1. INTRODUCTION

Determining the stage of decay (SOD) is a vital and common task in forensic anthropology. Knowing the degree of decomposition can reduce the potential pool of decedents that the human remains can belong to and help identify the deceased in human remain cases [1, 2]. Presently, establishing the SOD is done manually by trained experts via a thorough analysis of the corpse (e.g., soft tissue decomposition, skin discoloration, bone exposure, etc.), surrounding environment and climate, and presence of particular animals/insects. To the best of our knowledge, there is no research in the areas of computer vision for SOD estimation. We aim to make progress in this domain by attempting to build a supervised vision system to model human decay utilizing a large image collection collected over the past 10 years at the Forensic Anthropology Center at the University of Tennessee, Knoxville, which is home to the Anthropology Research Facility (ARF), also known as the "body farm", an outdoor decomposition laboratory.

Manually labeling the entire image collection, particularly because domain expertise is required, is prohibitive from both a resource and time-perspective. As is common for domain-specific datasets, the ARF image collection contains external (or exogenous) image attributes, such as the donor ID and the date contained in the image filename. Furthermore, since the protocol requires photos of distinct body parts, it is known that the internal (or endogenous) image attributes, apart from the SOD, include particular body parts. While the exogenous attributes are explicitly associated with each image, the endogenous attributes are known only implicitly, that is, from the general protocol used to do the photography. As such, endogenous image attributes, in order to be used, need to be obtained or derived via some sort of image classification technique just as the SOD. Besides the exogenous and endogenous image attributes, we also have basic domain knowledge concerning human decomposition. The salient aspect of the domain knowledge is that the decomposition varies across subjects (or donors) and body parts and is gradual. In other words, for the photos taken in the same session (or date) of the same subject, the same body part should have exactly the same SOD. The key idea of our label propagation method was to leverage the exogenous and endogenous image attributes in conjunction with domain knowledge to produce images similar to each image labeled by a forensic expert. The contribution of this work are summarized as follows:

1. A supervised vision system to perform SOD classification, utilizing a large human decomposition image collection, is presented and evaluated.
2. To reduce the need for costly and time-consuming manual labeling, a domain-aware label propagation method that leverages the exogenous and endogenous image attributes was developed. It is shown that our vision system using the proposed label propagation method provides improved SOD classification performance.

2. PROPOSED VISION SYSTEM

This section presents the developed vision system (see Figure 1) we called SOD-VS for Stage of Decay Vision System. The key parts of SOD-VS are (1) data preparation, which involves manual data labeling, label propagation using the proposed method, and splitting data into train, validation, and test sets, and (2) a

two-step transfer learning process to build and evaluate the SOD classification models. The SOD-VS implementation details and experimental settings are also presented.

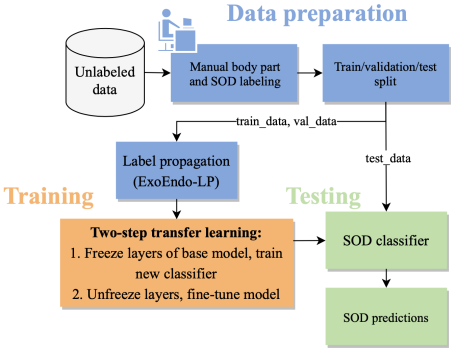

Fig. 1: Overview of the developed SOD-VS with the proposed label propagation method.

2.1. The human decomposition dataset

The dataset includes images of decomposing corpses donated to the ARF. The images were taken by forensic experts at non-uniform intervals (one or more days apart) from various angles and of different body parts to show the different stages and areas of human decomposition. The dataset spans from 2011 to 2022, and includes over 1.5 million images from over 800 donors.

2.2. Manual data annotation

The quality and size of the training data highly affects a model's performance and generalizability. That is, the more representative and diverse the training data is, the more generalizable a model is likely to be. In the case of a temporal dataset, such as images documenting human decomposition in which the subjects' appearance changes over time, it is important to sample the training data in such a way that it includes the dataset's characteristics. In our case, for instance, images depicting all possible decomposition stages should be included in the training data. Therefore, we randomly selected a small sample of subjects (or donors) and then a subset of images, of different body parts, taken from them over time (i.e., from when they first started to decay until fully decomposed) instead of randomly selecting images from the entire human decomposition dataset. This resulted in a subset of 4731 images of multiple body parts across different stages of decay, which were labeled by a forensic expert with a body part label (including head, torso, arm, hand, leg, or foot) and a decay score (DS) label, ranging from 1 (fresh) to 13 (fully skeletonized), using a modification of Galloway's [3] human decomposition scoring method to indicate the amount of decay present. The labeled dataset was named SHD: Stages of Human Decomposition. Figure 2 shows sample images of SHD. Since the aim is to predict the stage or level of decay rather than the DS itself, each image's DS label was mapped into DS intervals, carefully defined by the forensic experts, representing a certain SOD. As advised by the forensic experts, 3-4 stages are commonly used to describe the human decay process. Therefore, two versions of SHD were created by mapping each image's DS label into each of the following two sets of DS intervals: (1) 1-4 (fresh), 5-9 (active decay), and 10-13 (skeletonization), and (2) 1-4 (fresh), 5-7 (early decay), 8-9 (advanced decay), and 10-13 (skeletonization). The mapped versions of SHD, named SHD-3 (for 3 SOD classes) and SHD-4 (for 4 SOD classes), along with their DS groups (or intervals), class names, and frequencies (N) are shown in Table 1.

(a) No decay (BS=1). (b) Fully decayed (BS=13).

Fig. 2: Example images of the labeled dataset. The images show the same donor at different stages of decay. The image resolutions vary from 2400 × 1600 up to 4900 × 3200. Images are blurred to make them less disturbing to sensitive viewers.

Table 1: The SHD-3 and SHD-4 datasets along with their DS groups, class names, and class frequencies (N). Note, the class active decay = early + advanced decay.

Dataset	DS group	Class name	N
SHD-3	1-4	fresh	1479
	5-9	active decay	2053
	10-13	skeletonization	1199
SHD-4	1-4	fresh	1479
	5-7	early decay	1095
	8-9	advanced decay	958
	10-13	skeletonization	1199

2.3. Proposed label propagation method

The developed label propagation method leverages exogenous attributes, including the donor ID and timestamp (i.e., the date the image was taken), and endogenous attributes, including the body part shown in the image, to propagate the SOD labels of the labeled samples to the unlabeled samples. The image donor ID and timestamp are known attributes contained in the image filename, but the body part is unknown and is derived prior to label propagation using an existing body part classifier. This algorithm was named ExoEndo-LP for Exogenous-Endogenous-Based Label Propagation. ExoEndo-LP accepts the parameters, (1) id (image donor ID), (2) $time$ (image timestamp) and (3) bp (image body part), indicating the exogenous and endogenous attributes the algorithm considers during label propagation, which align with the attributes considered by the forensic experts during manual labeling. Note, the id parameter is a required parameter, while the $time$ and bp are optional parameters. Specifically, ExoEndo-LP works as follows: let $I = \{(x_i, (b_i, s_i), d_i, t_i)\}_{i=1}^{n}$ be a manually labeled dataset with n images, where (b_i, s_i) are the body part and SOD labels, respectively, d_i the donor ID,

Table 2: Performance comparison of the baseline method, LPA [4], and the proposed method, ExoEndo-LP, when applied to the train data (n=3312). Reported are average runtimes (in seconds) and the number of additionally labeled (NoL) train samples using label propagation.

Method	Runtime (s)	NoL
LPA [4]	32	14067
ExoEndo-LP($id, time$)	16	6034
ExoEndo-LP($id, time, bp$)	20	2250

and t_i the image timestamp for the i^{th} labeled image, x_i. If ExoEndo-LP($id, time$), meaning label propagation is subject and time-specific, then given a labeled image, $(x_i, (b_i, s_i), d_i, t_i) \in I$, all images $m \notin I$ of donor, d_i, with the timestamp, t_i, are found and denoted as the set $S_i = \{(x_j, d_i, t_i)\}_{j=1}^{m}$. Then, the SOD label, s_i, of the labeled image, x_i, is propagated throughout S_i resulting in the labeled set, $L_i = \{(x_j, s_i, d_i, t_i)\}_{j=1}^{m}$. Else if ExoEndo-LP($id, time, bp$), meaning label propagation is subject, time, and body part-specific, then given a labeled image, $(x_i, (b_i, s_i), d_i, t_i) \in I$, all images $m \notin I$ of donor, d_i, with the timestamp, t_i, are found and denoted as the set $S_i = \{(x_j, d_i, t_i)\}_{j=1}^{m}$. Since the *bodypart* parameter was passed in addition to the id and $time$ parameters, only images in S_i of the same body part as the labeled image, x_i, will be labeled with s_i. Therefore, a previously developed body part classifier (with a test macro-averaged F1 score of 0.93 and a test accuracy of 0.94) was used to automate this task by classifying each image in S_i into a body part (see Section 2.2 for body part labels), resulting in the set of l images denoted as $S_{i+} = \{(x_k, \hat{b_k}, d_i, t_i)\}_{k=1}^{l}$ where $\hat{b_k}$ is the predicted body part such that $\hat{b_k} = b_i$ for the k^{th} image, x_k. Then, the SOD label, s_i, of the labeled image, x_i, is propagated throughout S_{i+}, resulting in the labeled set, $L_i = \{(x_k, (b_i, s_i), d_i, t_i)\}_{k=1}^{l}$. In either case, the labeled set L_i gets added to a final set of labeled images, $I_+ = I \cup \{L_i\}_{i=1}^{n}$.

To evaluate the performance of the proposed ExoEndo-LP, the standard graph-based label propagation algorithm, (LPA) [4], was used as a baseline method. Unlike ExoEndo-LP, LPA [4] does not consider any image attributes, but is based on the assumption that closer data points tend to have similar class labels. Specifically, at every iteration of propagation, each unlabeled sample is labeled with the label that majority of its neighbors belong to or to a random selection when there is no single maximum. LPA [4] reaches convergence when each unlabeled sample has the majority label of its neighbors, or the predefined maximum number of iterations is achieved. To reduce the size of each labeled and unlabeled image, a previously trained classifier, excluding the last fully-connected and soft max layers, was used. Specifically, the output vectors of length 256 were used as the image vector representations to which LPA [4] was then applied.

2.4. SOD classification

To build the SOD classifiers, transfer learning was applied, which aims to produce effective models by leveraging and exploiting previously acquired knowledge [5]. In particular, three CNN architectures, including ResNet50 [6], Inception V3 [7], and Xception [8], pre-trained on the ImageNet [9] dataset, were trained using the following two-step transfer learning process: (1) freeze all pre-trained convolutional layers of the base model, and train newly added classifier layers and (2) unfreeze all layers, and fine-tune model end-to-end on the human decomposition dataset with a low learning rate. The newly added classifier layers consisted of five layers, including one global average pooling layer, one drop-out of 0.3 layer, two fully-connected layers with 128 and 64 nodes, respectively, and one final softmax layer with the number of nodes equal to the number of classes to perform multi-class classification. In addition, to increase diversity during training, a data augmentation layer, which performed image flipping (horizontal and vertical) and rotating, was added after the input layer. Adaptive Momentum Estimation (Adam) was used as the optimizer with a learning rate of 0.001 and 0.0001 for the first and second step in the two-step transfer learning process, respectively. Note, training a model from scratch and transfer learning without freezing the base model, and freezing only a certain number of base model layers followed by fine-tuning was also tested, however, model performances were drastically improved using the proposed two-step transfer learning process.

2.5. Implementation and experimental settings

The developed SOD-VS was implemented using Keras, TensorFlow, and Python, and MongoDB as the database. The SOD-VS was built and evaluated on the manually labeled datasets, SHD-3 and SHD-4, using the proposed ExoEndo-LP and baseline, LPA [4], methods to perform label propagation. Note, label propagation was only performed on the train and validation data, but not on the test data to ensure equal classification evaluation. To evaluate the proposed ExoEndo-LP, ExoEndo-LP($id, time$) and ExoEndo-LP($is, time, bp$) were tested. The baseline method, LPA [4], was implemented using 1000 maximum iterations and a varying number of neighbors (i.e., 5, 10, and 20). Therefore, a total of 30 experiments (2 datasets $\times$ (3 LPA [4] + 2 ExoEndo-LP) $\times$ three CNN architectures) were conducted. To train and evaluate the SOD classifiers, each labeled dataset (SHD-3 and SHD-4) was split into a train, validation, and test set using a ratio of 70:10:20, resulting in 3312 train samples, 473 validation samples, and 946 test samples. All images were resized to 224x224 and 299x299 depending on the model architecture. All models were trained on a single TeslaV100-SXM2 GPU with 32GB memory. The batch size was set to 32 and the number of epochs to 200 with early stopping set to 20 epochs (monitoring the validation loss) to avoid over-fitting on the training set. To evaluate the SOD classifiers, the per-class precision and recall, and macro-averaged F1 score (mF1) on the test data were calculated.

3. EXPERIMENTAL RESULTS AND DISCUSSION

The average runtimes and number of additionally labeled samples after the baseline method, LPA [4], and the proposed ExoEndo-LP were performed on the train data are reported in Table 2, which shows that the proposed ExoEndo-LP is up to twice as fast as the baseline method, LPA [4]. The baseline method, LPA [4], labeled

Method	Model	Per-Class Precision (P) & Recall (R)								
		fresh		early/active		advanced		skeletonization		
		P	R	P	R	P	R	P	R	mF1
SOD-VS +LPA [4]	ResNet-50	0.0	0.0	0.603	0.8	–	–	0.703	0.629	0.45
	Inception-V3	0.713	0.682	0.851	0.804	–	–	0.827	0.853	0.79
	Xception	0.7	0.554	0.819	0.844	–	–	0.834	0.844	0.763
SOD-VS +ExoEndo-LP($id, time$)	ResNet-50	0.77	0.555	0.396	0.22	–	–	0.551	0.91	0.54
	Inception-V3	0.869	0.844	0.723	0.863	–	–	0.824	0.861	**0.83**
	Xception	0.814	0.85	0.718	0.72	–	–	0.812	0.778	0.78
SOD-VS +ExoEndo-LP($id, time, bp$)	ResNet-50	0.797	0.441	0.522	0.911	–	–	0.771	0.157	0.497
	Inception-V3	0.927	0.844	0.813	0.906	–	–	0.892	0.809	**0.863**
	Xception	0.86	0.944	0.869	0.81	–	–	0.807	0.817	0.85
SOD-VS +LPA [4]	ResNet-50	0.5	0.068	0.67	0.324	0.0	0.0	0.568	0.944	0.318
	Inception-V3	0.757	0.818	0.701	0.64	0.627	0.576	0.709	0.629	0.683
	Xception	0.767	0.659	0.81	0.83	0.567	0.222	0.96	0.902	0.695
SOD-VS +ExoEndo-LP($id, time$)	ResNet-50	0.707	0.691	0.294	0.12	0.0	0.0	0.54	0.9	0.388
	Inception-V3	0.821	0.922	0.64	0.732	0.705	0.505	0.876	0.779	0.74
	Xception	0.91	0.954	0.696	0.771	0.71	0.468	0.911	0.884	**0.78**
SOD-VS +ExoEndo-LP($id, time, bp$)	ResNet-50	0.567	0.801	0.349	0.494	0.288	0.108	0.631	0.277	0.403
	Inception-V3	0.844	0.916	0.706	0.705	0.642	0.675	0.923	0.766	0.773
	Xception	0.897	0.908	0.734	0.8	0.75	0.619	0.851	0.872	**0.803**

Table 3: SOD-VS with the baseline method, LPA [4] (neighbors=5), and the proposed ExoEndo-LP evaluated on the SHD-3 test data (top three) and SHD-4 test data (bottom three).

more samples than the proposed ExoEndo-LP, which aligns with the fact that since ExoEndo-LP considers images attributes, it is more restrictive, meaning less candidates are considered during labeling, resulting in less labeled images than with LPA [4] that labels each unlabeled image based on the majority class of its neighbors. Although, ExoEndo-LP produced a smaller labeled train set, higher SOD classification performance was achieved as shown by Table 3. This suggests that our method produced synthetic labels of quality more similar to that of manual labels produced by the forensic experts, instead of more numerous but, presumably, less accurate synthetic labels produced by the traditional label propagation. Table 3 shows that the SOD-VS with the proposed ExoEndo-LP outperforms the SOD-VS with the baseline method, LPA [4], when built on the SHD-3 (3 SOD classes) and SHD-4 (4 SOD classes) datasets. Note, SOD-VS with LPA [4] performed best using 5 neighbors so only these results are reported in Table 3. The highest SOD classification performance on the SOD-3 test data (overall the highest performer) with a mF1 score of 0.863 was achieved by SOD-VS with ExoEndo-LP($id, time, bp$) and Inception-V3, and for the SOD-4 test data with a mF1 score of 0.803 was achieved by SOD-VS with ExoEndo-LP($id, time, bp$) and Xception. Overall, classification results were higher when three SOD classes were used, that is, when early and advanced decay were combined which according to the forensic experts are the most difficult to differentiate between. Indeed, "advanced" decay had the lowest recall of only 0.62 for the ExoEndo-LP($id, time, bp$) using Xception. ResNet-50 was outperformed in all experiments, which could be due to its deep architecture requiring longer training time than more shallow architectures, such as Inception-V3 and Xception.

4. CONCLUSION

This paper presents a novel vision system that utilizes a large human decomposition image dataset to perform SOD classification. The system uses a label propagation method that leverages the exogenous and endogenous image attributes with the constraints based on domain knowledge to improve classification performance (and reduce the manual labeling efforts). The approach is general in the sense that image collections in other domains tend to have exogenous and endogenous attributes that could be used in a manner similar to the one employed in this work.

The improvements from the basic label propagation make the classification sufficiently accurate for a number of downstream applications, such as Post-Mortem Interval estimation. Future work will focus on improving the classification performances of the early and advanced decay classes since all CNN architectures struggled to distinguish the two to improve the accuracy of the four SOD classes case. Simultaneous estimation of multiple endogenous attributes will also be considered as the estimates of SD may be used to improve identification of body parts as well.

We hope this work can help progress the research of forensic anthropology in conjunction with computer vision and be used in other domains where, similar to human decay, a large degree of variation exists and human labeling is limited from both a time and resource-perspective.

5. REFERENCES

[1] Mary S Megyesi, Stephen P Nawrocki, and Neal H Haskell, "Using accumulated degree-days to estimate the postmortem interval from decomposed human remains," *Journal of Forensic Science*, vol. 50, no. 3, pp. 618–626, 2005.

[2] HT Gelderman, L Boer, Tatjana Naujocks, ACM IJzermans, and WLJM Duijst, "The development of a post-mortem interval estimation for human remains found on land in the netherlands," *International journal of legal medicine*, vol. 132, no. 3, pp. 863–873, 2018.

[3] Alison Galloway, Walter H Birkby, Allen M Jones, Thomas E Henry, and Bruce O Parks, "Decay rates of human remains in an arid environment," *Journal of Forensic Science*, vol. 34, no. 3, pp. 607–616, 1989.

[4] Usha Nandini Raghavan, Réka Albert, and Soundar Kumara, "Near linear time algorithm to detect community structures in large-scale networks," *Physical review E*, vol. 76, no. 3, pp. 036106, 2007.

[5] Sinno Jialin Pan and Qiang Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[7] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[8] François Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.

[9] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *IEEE conference on computer vision and pattern recognition*, 2009, pp. 248–255.