

Person Authentication using Head Images

Aakarsh Malhotra, Richa Singh, Mayank Vatsa
IIIT-Delhi, India
{aakarshm, rsingh, mayank}@iiitd.ac.in

Vishal M. Patel
Rutgers University, USA
vishal.m.patel@rutgers.edu

Abstract

In many surveillance applications, the cameras are placed at overhead heights for human identification. In such real-world scenarios, the person of interest might be walking away from the camera and the only information available is “image of the person’s head”. In this research, we investigate the usage of head images for person recognition and propose it as a soft-biometric modality. With its viability for human recognition, application of head images can also be extended with other face recognition algorithms for surveillance. We propose a head image database pertaining to 103 subjects with more than 600 images. In addition to the database, we propose a framework for head image-based person verification. As a pre-processing stage, the framework includes evaluation of two segmentation algorithms. We also perform benchmarking evaluations of various texture, key-point, and learning-based representation algorithms and establish the baseline results. The experiments suggest that head images can be effectively used to ascertain human identity and the availability of this database could pave further research in this field.

1. Introduction

With the increase in crime, surveillance systems [8] have gained significant attention in the past few decades. Such surveillance systems typically work under unconstrained environments and encounter many challenges. These challenges arise from either environmental factors or due to lack of cooperation from the user. For instance, environmental challenges include illumination variation, background noise, and blurriness. On the other hand, challenges such as user walking away from the camera, partial face, and occlusion occur due to covert acquisition in surveillance systems. In security systems, users may also try to spoof the system by using a mask to cover the frontal side of their face. In such cases, soft biometrics [4], [15], [17] act as an alternative to aid the performance of recognition systems. Soft biometric traits include physical and behavioral characteristics [17]. However, in some cases, there might not be any



Figure 1: Sample images illustrating the usage of head images for recognition in surveillance footage* (cases where no facial information is visible).

*Source: <https://tinyurl.com/yamdf4o7>, <https://tinyurl.com/y926f9ly>

biometric information visible. One such scenario is surveillance, as shown in Figure 1. The only information visible is the head region of the human body.

The head image is an image of the head region of the human body, acquired from an overhead camera. In the acquired image, the face may or may not be visible. These images contain important features which can be utilized for human recognition. The head-top hair is often used by people to recognize their known ones. These hairs have distinctive characteristics such as style, color, straightness or curliness, and hair parting. Building upon such characteris-

tics, a biometric system can potentially determine a human identity when the frontal view of the face is not visible or clear. As seen in Figure 1(a), the suspects for shoplifting refrained themselves from looking into the only CCTV camera of the shop. Similarly, in Figure 1(b), the suspect who is carrying ammunition is walking away from the camera. The footage in this case contains the head region only. Hence, to utilize distinctive traits contained in the images of head, we ask a simple question in this research: can head images be employed for human recognition?

1.1. Literature Review

In the past two decades, a significant amount of research has been performed for surveillance-based recognition. However, only a few approaches include the usage of facial and head hair for recognition purposes. These methods can broadly be classified into two categories, head images only and facial images. In head images only, the biometric system takes into account only the information available on head. This might include the color of hair, shape of the head, and hairstyle. On the other hand, the facial approaches take into account facial hair (such as beard, mustaches, and front-view hair). It may also include biometric traits of the face such as shape and facial features.

The existing research on head images is primarily focused on utilizing color and textural information for feature extraction for head images. However, a significant limitation in each of these work is the small size of the database. The largest database utilized in such approaches include head images from only 30 subjects. In 2000, Cohen et al. [2] reported an accuracy of 96% on a dataset of 12 subjects using decision trees. The segmentation is performed by background subtraction using chrominance components. Texture features such as directionality, contrast, LBPHS, and coarseness are considered. Authors also took into account additional color information of clothing, body, and hair color.

Nakajima and Sasaki [13] reported 86.4% recognition rate using 33 head images from eleven subjects acquired using a thermo camera. Their algorithm involved head detection and calculating 2D-FFT, which is then matched with the gallery. In 2005, authors extended their work [12], where they obtained 2D-DCT instead of 2D-FFT. Using the same dataset, authors increased the system performance to 100%. Aradhye et al. [1] characterized hair using line segment based features and pixel based properties. These features included: (i) macro-texture (orientation and length), (ii) shape, (iii) color, and (iv) features computed using MR-RISAR model [11]. The matching is performed by calculating the likelihood of a hair patch belonging to a person. For multiple enrollments, the similarity is based only on line-based features. The pixel-based features are considered in case of a tie. Authors reported a performance of 77% and

92% for single enrollment image and multiple enrollment image matching respectively.

The other stream of research considers facial hair and partial face (if available) for recognition. The methods primarily rely on using color space information and classifying hair into different styles. Such approaches were first introduced by Jia [7] in 1992. Zhang et al. [24] performed head detection on video surveillance using XYZ and HSV color space. They modeled head shape by calculating the likelihood for each color model. Wang and Ai [21] classified hair into seven hairstyles by finding the most informative patches using the RankBoost algorithm. In 2013, Dass et al. [3] discovered hairstyle from frontal face images for recognition. They considered darker pixels as hair and clustered them. The clustered information is combined with background and face-skin masks to classify structure into five hairstyles. In 2014, Wang et al. [22] performed human hair segmentation and length detection of hair using histogram analysis and K-means clustering. Recently, Proença and Neves [16] used a multi layered Markov Random Fields (MRF) for segmentation and classification of details present on the face. Using 81 features (R, G, B, H, S, V, Y, Cb, Cr intensities, and LBP) for each pixel, three supervised classifiers gave the posterior probability for each class, namely, hair, skin, and background.

1.2. Contributions

The literature review highlights that the research in the domain of head images only methods is very limited. While facial approaches have shown results on larger datasets, the approaches related to head images have evaluated their results only on small datasets with a maximum of 30 subjects. With 30 subjects, it is hard to make a generalized claim on the effectiveness of head images as a suitable soft-biometric. Hence, to answer the question of using head images for recognition, a larger database is required. Therefore, we develop Head-image Soft Biometric Database (HSBD) as a part of this study and make it available for the biometric research community. The HSBD proposed in this study is the first ever database which has no facial information and has a well defined protocol. To promote research in this field, the database will also be made publicly available for research¹. The database consists of more than 600 images from 103 subjects. The head images contain no facial information, thereby making the recognition problem exciting and challenging. This research also establishes the baseline results using hand-crafted and representation-based features on HSBD. These experiments fall into the category of head image only approach since the HSBD contains no facial information. The experiments show that head images can be utilized as a soft biometric in surveillance scenarios with over 90% accuracy at 10% FAR

¹<http://iab-rubric.org/resources/HSBD.html>



Figure 2: Sample images of two subjects from the proposed Head-images Soft Biometric Database (HSBD). Images are captured in different sessions with varying hairstyles.

using deep learning-based segmentation and feature extraction.

The remaining paper is organized as follows: Section 2 elaborates about the database collected, Section 3 discusses the proposed matching framework and Section 4 summarizes the baseline results on the proposed dataset.

2. Head-images Soft Biometric Database

To the best of our knowledge, there is no publicly available database which captures head images. In this research, we present the Head-images Soft Biometric Database (HSBD)² to promote research in the field of human recognition. For the collection of database, the approval from ethics board has been obtained, and for minor subjects, the approval is obtained from school administration. Along with this, for adult participants, their consent has also been taken.

2.1. Acquisition

All the head images in the database have been acquired using DSLR cameras. The images are captured in natural daytime lighting without the use of flash. The cameras used in the acquisition include Nikon D90 DSLR camera with 12.3 MP resolution and Canon EOS 70D with 20 MP resolution. For both the cameras, autofocus is kept ON to acquire images at optimal quality settings.

While the head images are acquired, it is ensured that the participant does not wear any caps or head scarfs to cover their head. However, they are allowed to use hair clips and bands to tie their hair. This procedure is performed to ensure that we capture hair in their head images and properties of

their hair can be utilized for recognition. Figure 2 illustrates sample images from the proposed HSB database.

2.2. Database Statistics

The entire database consists of head images acquired from participants of age ranging from 2 years to 38 years. Each of the participants belonged to Indian nationality and had black or dark brown hair. The proposed database consists of 606 images from 103 subjects. Out of these 103 subjects, there are 55 male participants and 48 female participants. For each subject, a minimum of 3 images and a maximum of 7 images have been acquired. For some participants, there are angular variations in the acquired head image. The data from each participant is indexed using a four-digit random identifier, and all images are stored in JPEG format. A summary of the dataset is provided in Table 1. In addition to these images, a supplementary set of 18 post haircut images corresponding to 5 subjects is also included in the database.

To establish the exact location of head in the acquired image, manual annotation of each of the database images is also provided. Since the head image is axis aligned, we locate the head by $xmin$, $xmax$, $ymin$, and $ymax$ points in the image. These points can be used to evaluate the segmentation performance. Along with the database, these annotations will also be released.

2.3. Database Protocol

The HSB database consists of 103 subjects in total. The database is split according to identities of the subjects. All images from 52 subjects are used as the training set and the remaining 51 subjects comprise the testing set. While per-

²Database link: <http://iab-rubric.org/resources/HSBD.html>

Table 1: Characteristics of the proposed HSB Database.

No. of samples	No. of Subjects	No. of Images
Three	5	15
Four	14	56
Five	16	80
Six	21	126
Seven	47	329
Total	103	606

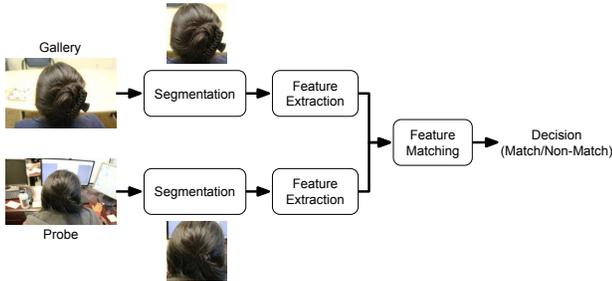


Figure 3: Proposed head images matching framework.

forming training and testing, the two tidiest head images (as per visual appearance) are considered as the gallery. The remaining samples of the subject (ranging from 1 to 5) are used as query images. The mentioned split is used for performing verification experiments in the proposed framework.

3. Proposed Head Images Matching Framework

Using the protocol mentioned in the previous section, we propose a head image matching framework, summarized in Figure 3. In this research, we have evaluated the performance of multiple segmentation, feature extraction, and matching algorithms to establish the baseline results. Each of these steps is explained in detail in the following sections.

3.1. Segmentation

To find the exact region of interest (ROI), we need to segregate the region where only the head is present. This ensures that no background object is matched across images and only traits unique to the subject identity are matched. This is achieved using segmentation from two different procedures as described below.

3.1.1 Image processing-based segmentation

The following 3-step pipeline is used for segmentation.

- **Step 1:** The input image is binarized and then inverted to select darker regions. Thus, the “white pixels” represent the darker regions (hence, dark-haired regions).



Figure 4: Sample segmentation output using image processing-based segmentation. First row shows input images while the second row presents the segmentation output.

Assuming that the head is present in the center of the image, 10% of the image is discarded from left, top, right, and bottom of the image. Then the largest connected component is found and all other components are removed. To fill up the empty spaces in the connected component, hole filling operation is performed. The boundary of the largest component is smoothed using dilation operation.

- **Step 2:** Some background objects still remain in the largest connected component. To remove background noise, we begin from the center of each row. We extend towards the left boundary of the image, looking for the first transition from a white to a black pixel. Once we find first such occurrence, all pixels are made black beyond that point. The same procedure is repeated while extending towards the right boundary from the center pixel of the row.
- **Step 3:** To remove small components that are generated due to the previous step, we again find the largest connected components and remove all other components.

Output Image: Using the output from Step 3, we map the segmentation boundary to the original image. ROI is found by locating the extreme white pixel at each side and drawing a bounding box to obtain the segmented image. The segmentation performance is reported in Table 2 and Figure 4 illustrates some sample segmented head images.

3.1.2 Segmentation using FRCNN

The second technique for segmentation utilizes Faster-RCNN [18]. Head region segmentation is considered as a two-class problem with the classes being head and background. For images acquired in different resolutions, the Faster-RCNN network can detect objects with their classification scores.

The architecture of the Faster-RCNN has convolutional neural network to first generate feature maps. Features

Table 2: Segmentation performance of the proposed segmentation techniques for head images.

Segmentation Method	SA (%)	FSA (%)	BSA (%)
Image Processing	91.54	89.48	92.80
FRCNN	96.59	94.63	97.59

maps are generated from a ZF-Net [23] architecture and is then used by the Region Proposal Network (RPN) to generate region proposals. The other part is for object detection using Fast-RCNN. Each of the object proposals is converted to a fixed size feature map using region of interest pooling. Subsequently, the fixed size feature map is fed into fully connected layers (FCs) to obtain feature vector. Using two different layers, per-class bounding-box regression and class probabilities are obtained for every feature vector.

The experiments are performed using a model pre-trained on the Pascal VOC dataset [5]. The model is fine-tuned for two-class segmentation problem using the training partition from HSB Database. To further enhance the fine-tuning process, the data is augmented by translation, mirroring, blurring, and changing pixel intensities of the original 317 training images. The bounding-box returned by the model on test images is used to crop and segment head regions. Figure 5(c) shows samples of the segmentation using Faster-RCNN model.

The results of the proposed segmentation algorithm are evaluated in terms of three metrics: Segmentation accuracy, foreground segmentation accuracy, and background segmentation accuracy. Segmentation accuracy is defined as follows:

$$SA = \frac{CCP}{TP} \quad (1)$$

where, CCP is the count of correctly classified pixels, and TP is the total number of pixels. The foreground segmentation accuracy can be evaluated as follows:

$$FSA = \frac{CCFP}{TFP} \quad (2)$$

where, CCFP is the count of correctly classified foreground pixels, and TFP is the total number of foreground pixels. Similarly, background segmentation accuracy can be evaluated as follows:

$$BSA = \frac{CCBP}{TBP} \quad (3)$$

where, CCBP is the number of correctly classified pixels and TBP are the total number of background pixels. The segmentation performance of the algorithm can be seen in Table 2. Once we obtain the segmented image, the next step is feature extraction.

3.2. Feature Extraction

In head images, a slight variation in hairstyle can either reduce the inter-class variations or increase intra-class sim-

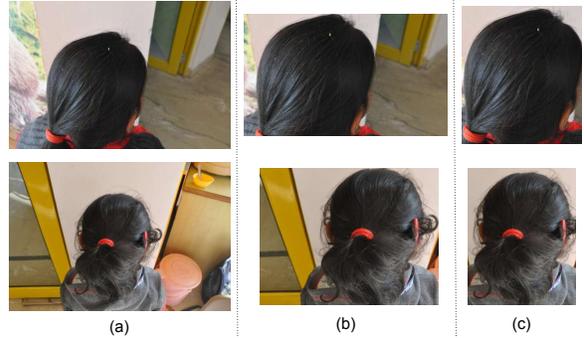


Figure 5: Sample illustrations of (a) original image, (b) Image processing based segmentation, and (c) Faster-RCNN segmentation output.

ilarities. Therefore, we need to explore various representation extraction/learning and classification techniques which can handle variations. This research explores various hand-crafted features and learning-based representations for head image based person recognition. The hand-crafted features utilized include Dense SIFT (DSIFT) and Local Binary Patterns (LBPHS). The intuition for using these features is to analyze both key-point based representation and texture representation. Learning-based representations considered in these experiments include dictionary representations, VGGFace, and ResNet50. Different experiments are tabulated in Table 3 and the list of features used in this research are described below.

- **Hand-Crafted Features [10] [14]:** In this experiment, the DSIFT [10] and LBPHS [14] features are extracted from the gallery and probe images. Both of these features are obtained for each color channel of the RGB color space and the corresponding grayscale image. For each of these cases, the segmented image is down-scaled to a fixed size of 224×224 . For the DSIFT features, a spatial bin size of 12 and a step size of 32 pixels is chosen. The resultant feature is of size 128×36 . For LBPHS, the cell size is kept as 32×32 , using which feature representation of size 58×49 is obtained. These values are selected experimentally to yield the best performance.
- **Dictionary [9]:** In Experiment 2(a) and 4(a), dictionary atoms are learned from training head images corresponding to 52 subjects. Using these atoms, feature representation is obtained for the downsampled segmented grayscale image (224×224) and RGB image ($224 \times 224 \times 3$). A dictionary representation $D_{N \times K}$ is

obtained using the following optimization [20]:

$$D_{opt}, W_{opt} = \arg \min_{D, W} \sum_{l=1}^L \|w_l\|_p + \gamma \|X - DW\|^2$$

$$\text{and } \|d_i\| = 1$$
(4)

Here, N is the size of the image after vectorization, K is the number of atoms used in creating the dictionary, L is the number of samples used for training, $X_{N \times L}$ is the training matrix, $p = 1$, and $W_{K \times L} = [w_1, w_2, \dots, w_L]$ is the coefficient matrix. We have used $K = 100$ dictionary atoms. The aim of the optimization is to minimize representation error $R = X - DW$, while fulfilling sparseness criterion for W . Once atoms are learned, the feature representation for each test head image is obtained.

- **VGGFace [19]:** In Experiments 2(b) and 4(b), the input segmented RGB image is downsampled to 224×224 and the features from pre-trained VGGFace architecture [19] are obtained. A global average pooling layer is applied after the last convolution layer to obtain a feature vector of length 512.
- **ResNet50 [6]:** In Experiment 2(c) and 4(c), ResNet50 architecture [6] is used to extract features for the head images. Similar to the other experiments, the input segmented RGB image is downsampled to 224×224 . The weights of ResNet50 are initialized by the weights of the model trained on the ImageNet database. The top layer of Softmax is removed and a feature vector of length 2048 is obtained.

3.3. Feature Matching

Once the features are extracted, the features of the probe images b are matched with the gallery images a . The comparison between the feature templates is performed using both cosine similarity and Euclidean distance metric. We observed cosine similarity between the gallery representation a and the probe representation b yields better performance. Therefore, the results are reported using cosine similarity only.

4. Results and Analysis

In section 3.1, we define the metrics for evaluating the segmentation performance of the proposed techniques. These metrics include Segmentation Accuracy (SA), Foreground Segmentation Accuracy (FSA), and Background Segmentation Accuracy (BSA). Using these metrics, we compare the performance of the two segmentation methodologies and their results are highlighted in Table 2. The first

one is based on image processing operations while the second technique utilizes FRCNN. FRCNN yields a high segmentation accuracy of 96.59%, highlighting the suitability of FRCNN for the application of head image recognition.

The experiments are performed with 5 times random subsampling based cross-validation and results of all the experiments are presented in Table 3. The corresponding receiver operating characteristic (ROC) curves for these experiments are shown in Figure 6 (Experiment 1 and 3) and Figure 7 (Experiment 2 and 4). The major conclusion which can be drawn from the results are as follows:

On evaluating texture features (LBPHS) and key point based features (DSIFT) for matching head images, it is observed that LBPHS features outperform DSIFT features by over 10% genuine accept rate (GAR) for all color channels. While hair orientation might be altered, the texture features can encode hairstyle better in head image recognition. Due to the better encoding of hairstyle, LBPHS has better discriminative capability compared to DSIFT in the problem of head image segmentation.

Representations extracted using trained dictionary perform poorly. On grayscale images, representations extracted from dictionary have a GAR of 24.28% compared to 62.97% by LBPHS at 1% FAR. There is a significant difference of over 38.6% at 1% FAR between GAR of dictionary representations and GAR from LBPHS features. This shows that representation derived from dictionaries matched using cosine similarity may not be a good framework for such a problem. Further, the performance of dictionary features, LBPHS and DSIFT features are lower on grayscale images compared to RGB.

Representations obtained from deep models give the most competitive results. Though VGGFace has a GAR of 83.35% at 10% FAR, the best results of all the representations are obtained using ResNet50. At 10% FAR, the GAR is 90.01% when using FRCNN based segmentation. It indicates that the weights learned using ImageNet in ResNet50 help in achieving better representation compared to VGGFace model. Since VGGFace learns the facial features, lack of such features in head-images degrades the performance of VGGFace based architecture. It is observed that ResNet50, VGGFace, and LBPHS features perform relatively better, each of them performing over 81% at 10% FAR. It can be attributed to the encoding power of ResNet50, VGGFace, and LBPHS features to encode line based feature [1] and textural features. Also, an early pre-processing step which segments out the head region is necessary. The segmentation technique utilized in the proposed framework is Faster-RCNN.

Even with such sophisticated architectures with ResNet50, the overall accuracies for head image recognition are quite low with the best Equal Error Rate (EER) of 11.19%. Additional experiments are performed with

Table 3: GAR of the proposed head image matching framework at various False Acceptance Rate (FAR).

Segmentation	Feature Type	Exp. No.	Feature	Color Channel	0.1% FAR	1% FAR	10% FAR
Image Processing	Hand-Crafted Features	Exp. 1(a)	DSIFT	Gray	16.77 ± 0.48	31.22 ± 0.67	54.03 ± 0.72
				R+G+B Fusion	17.02 ± 0.50	31.47 ± 0.64	54.28 ± 0.75
		Exp. 1(b)	LBPHS	Gray	30.25 ± 0.04	47.79 ± 0.86	69.84 ± 0.83
				R+G+B Fusion	30.67 ± 0.87	48.41 ± 0.74	70.61 ± 0.40
	Representation Learning	Exp. 2(a)	Dictionary	Gray	5.19 ± 2.05	20.82 ± 1.84	51.18 ± 0.92
				RGB	7.03 ± 2.76	18.59 ± 2.47	52.87 ± 1.41
Exp. 2(b)				VGGFace	RGB	23.10 ± 2.12	46.36 ± 2.49
Exp. 2(c)	ResNet50	RGB	31.34 ± 7.48	52.51 ± 2.88	83.87 ± 5.55		
FRCNN	Hand-Crafted Features	Exp. 3(a)	DSIFT	Gray	21.61 ± 2.34	41.93 ± 1.20	68.81 ± 1.38
				R+G+B Fusion	21.84 ± 2.01	42.33 ± 1.77	69.06 ± 1.40
		Exp. 3(b)	LBPHS	Gray	39.73 ± 5.07	61.95 ± 1.63	82.06 ± 4.12
				R+G+B Fusion	39.92 ± 6.32	62.97 ± 2.08	81.86 ± 3.08
	Representation Learning	Exp. 4(a)	Dictionary	Gray	6.51 ± 2.40	22.14 ± 2.19	58.20 ± 1.77
				RGB	6.93 ± 2.62	24.48 ± 2.47	62.26 ± 1.84
		Exp. 4(b)	VGGFace	RGB	28.15 ± 0.64	55.47 ± 2.09	83.35 ± 3.30
				Exp. 4(c)	ResNet50	RGB	42.96 ± 7.31

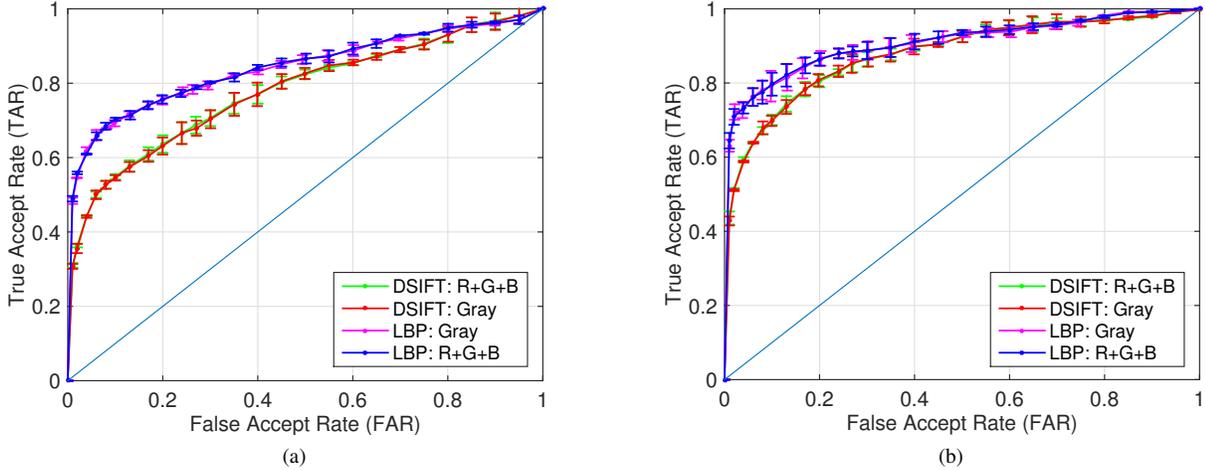


Figure 6: ROC curves for head image verification experiments. (a) Experiment 1: Hand-crafted features with image processing segmentation, (b) Experiment 3: Hand-crafted features with FRCNN segmentation.

18 pre-haircut and post-haircut images. ResNet50 yields verification accuracy of about 38%. The primary reason for low accuracy is large variations in pre-post haircut.

Unlike established biometric modalities such as fingerprints and face, the head images lack unique traits which are required for biometric recognition. However, since they provide a source of recognition in absence of more popularly used biometric modalities, it is our assertion that head images should be treated as a soft biometric modality with distinctive physical characteristics and researched upon further. It has an immense application for head images acquired using an overhead surveillance camera. However, various challenges such as hair fall, recoloring, changing hairstyle, and haircuts hinder the matching performance.

One such example is illustrated in Figure 8, where the participant got a haircut. The first image is used in the gallery and post-haircut head image is used as the probe image. Using features generated from ResNet50 model, the feature representation of the two head images failed to match. However, we would like to emphasize that head image is “soft” feature and therefore, such variations in accuracies are expected. We plan to investigate it more and in future, propose improved algorithm.

5. Conclusion

With increasing crime rates, human recognition and tracking in surveillance applications have been in focus. Though face recognition algorithms are becoming robust

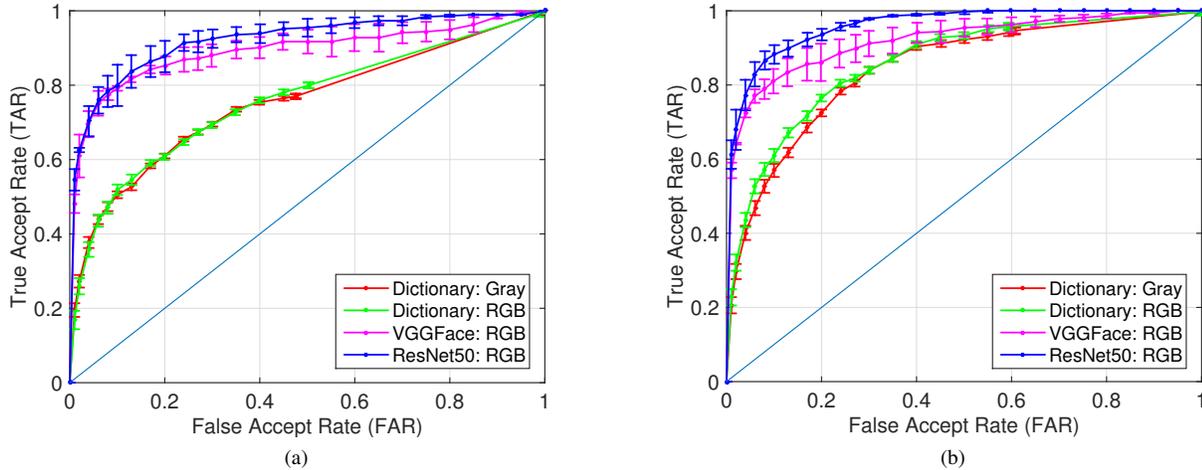


Figure 7: ROC curves for head image verification experiments. (a) Experiment 2: Representation learning with image processing segmentation, (b) Experiment 4: Representation learning with FRCNN segmentation.

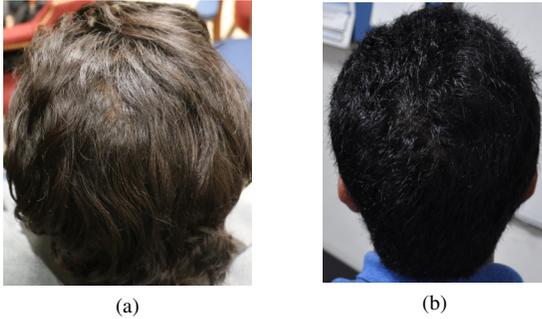


Figure 8: Illustration of a misclassified head image pair due to the effect of haircut. (a) Gallery image (with long hair) and (b) probe image (after haircut).

day by day, they fail to recognize and track a human if only the head portion is visible. In this study, we propose the first ever public database which includes head-images captured using a DSLR camera. The proposed Head-image Soft Biometric Database (HSBD) contains more than 600 images from 103 subjects with a diverse age range from 2 to 38 years. This research illustrates that use of head images as a soft-biometric can provide a decent genuine accept rate of around than 81% using LBPHS texture features and around 90% using deep models. With such a performance, the Head-image Soft Biometric Database (HSBD) has a significant scope for research in human tracking and recognition. In challenging scenarios such as surveillance, head images can also be utilized by fusing with face to improve the performance of the biometric system.

In head images, several challenges need to be addressed. While pose variation remains the foremost, other challenges

such as baldness, changes in hairstyle, and change in color would hinder the system performance. However, these challenges highlight that head image recognition requires significant attention and research efforts. With making this database publicly available for the research community, we hope that it would promote research on this vital topic.

6. Acknowledgment

This research is partially supported by Ministry of Electronics and Information Technology, India. We would like to acknowledge Mallika Agarwal for her assistance in creating the dataset. The authors would also like to thank all the volunteers for co-operating in the data collection process. Aakarsh Malhotra is partially supported by Visvesvaraya Ph.D. fellowship. Richa Singh and Mayank Vatsa are partly supported by the Infosys Center of Artificial Intelligence, IIT Delhi, India. Vishal M. Patel was supported by US Office of Naval Research (ONR) Grant YIP N00014-16-1-3134.

References

- [1] H. Aradhye, M. Fischler, R. Bolles, and G. Myers. Headprint–Person Reacquisition using Visual Features of Hair in Overhead Surveillance Video. In *International Conference on Audio-and Video-Based Biometric Person Authentication*, pages 879–890. Springer, 2005.
- [2] I. Cohen, A. Garg, and T. S. Huang. Vision-based Overhead View Person Recognition. In *International Conference on Pattern Recognition*, volume 1, pages 1119–1124. IEEE, 2000.
- [3] J. Dass, M. Sharma, E. Hassan, and H. Ghosh. A Density based method for Automatic Hairstyle Discovery and Recognition. In *National Conference on Computer Vision*,

- Pattern Recognition, Image Processing and Graphics*, pages 1–4. IEEE, 2013.
- [4] S. Denman, C. Fookes, A. Bialkowski, and S. Sridharan. Soft-biometrics: Unconstrained Authentication in a Surveillance Environment. In *Digital Image Computing: Techniques and Applications*, pages 196–203. IEEE, 2009.
- [5] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results. <http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html>.
- [6] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Computer Vision and Pattern Recognition*, pages 770–778. IEEE, 2016.
- [7] X. Jia and M. Nixon. Extending the Feature set for Automatic Face Recognition. In *International Conference on Image Processing and its Applications*, pages 155–158. IET, 1992.
- [8] L. W. Kille and M. Maximino. The effect of CCTV on public safety: Research roundup. <https://journalistsresource.org/studies/government/criminal-justice/surveillance-cameras-and-crime>, (accessed 2017-9-8).
- [9] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.
- [10] D. G. Lowe. Object recognition from local scale-invariant features. In *International Conference on Computer vision*, volume 2, pages 1150–1157. IEEE, 1999.
- [11] J. Mao and A. K. Jain. Texture Classification and Segmentation using Multiresolution Simultaneous Autoregressive Models. *Pattern Recognition*, 25(2):173–188, 1992.
- [12] K. Nakajima, A. Kamiya, and K. Sasaki. Personal Recognition Rate Improvement Using Head-Top Image and An Application to Walking Subject. In *International Conference of the Engineering in Medicine and Biology Society*, pages 3753–3755. IEEE, 2005.
- [13] K. Nakajima and K. Sasaki. Personal Recognition using Head-Top Image for Health-Monitoring System in the Home. In *International Conference of the Engineering in Medicine and Biology Society*, volume 2, pages 3147–3150. IEEE, 2004.
- [14] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.
- [15] U. Park and A. K. Jain. Face Matching and Retrieval using Soft Biometrics. *IEEE Transactions on Information Forensics and Security*, 5(3):406–415, 2010.
- [16] H. Proença and J. C. Neves. Soft Biometrics: Globally Coherent Solutions for Hair Segmentation and Style Recognition Based on Hierarchical MRFs. *IEEE Transactions on Information Forensics and Security*, 12(7):1637–1645, 2017.
- [17] D. Reid, S. Samangoei, C. Chen, M. Nixon, and A. Ross. Soft Biometrics for Surveillance: an Overview. *Machine Learning: Theory and Applications*. Elsevier, pages 327–352, 2013.
- [18] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*, pages 91–99, 2015.
- [19] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [20] K. Skretting. Dictionary Learning Tools for Matlab. <http://www.uis.no/karlsk/dle/>.
- [21] N. Wang and H. Ai. Hair Style Retrieval by Semantic Mapping on Informative Patches. In *Asian Conference on Pattern Recognition*, pages 110–114. IEEE, 2011.
- [22] Y. Wang, Z. Zhou, E. K. Teoh, and B. Su. Human Hair Segmentation and Length Detection for Human Appearance Model. In *International Conference on Pattern Recognition*, pages 450–454. IEEE, 2014.
- [23] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. In *European Conference on Computer Vision*, pages 818–833. Springer, 2014.
- [24] Z. Zhang, H. Gunes, and M. Piccardi. Head Detection for Video Surveillance Based on Categorical Hair and Skin Colour Models. In *International Conference on Image Processing*, pages 1137–1140. IEEE, 2009.