VTU-BEC-DB Multimodal Biometrics Database

The VTU-BEC-DB Multimodal Biometrics Database for Biometric research is constructed by collecting the three biometric traits viz. face, hand and speech/voice from 100 persons, including 36 males and 64 females with age range between 18 and 50 years. The database is collected at the, Department of Computer Science and Engineering at Basaveshwar Engineering College(BEC), Bagalkot affiliated to Visvesvaraya Technological University(VTU), Belagavi, Karnataka, India. VTU-BEC-DB comprises three sub-databases namely a face database (FaceDB), a hand database (HandDB) and voice corpus(VoiceDB) which are collected from 100 users/subjects. These sub-database can be used separately to test the uni-modal biometric systems by employing the proposed experimentation protocols. The brief description of the VTU-BEC-DB Multimodal Biometrics Database is given in Table 1.

Table 1: VTU-BEC-DB Multimodal Biometrics Database

Biometric Trait	Number of sample images/voiceprints per subject	Total number of images/voiceprints in the database/ corpus	Description
Face	30	3000	**15 samples** collected by adjusting the pixel resolution of the camera to 5 Megapixel and 10 Megapixel with different illuminations and pose variations.
Hand	24	2400	**12 samples** collected from left and right hand with illumination variations.
Voice	620	62000	**310 voiceprints** recorded in Kannada and English languages separately.

Purpose: The VTU-BEC-DB Multimodal Biometrics Database is meant to be used for research purposes and shall not be used nor included in commercial applications in any form (e.g., original files, encrypted files, files containing extracted features, etc).

Citation: All documents and papers that use the VTU-BEC-DB Multimodal Biometrics Database must acknowledge the use of the database by including the following reference.

Angadi Shanmukhappa A., Hatture Sanjeevakumar M., 2018," Multimodal Biometrics Database for Person Authentication : VTU-BEC-DB," Proceedings of the International Conference on Intelligent Computing and Sustainable System(ICICSS 2018), pp. 573-579, Coimbatore, India, 20-21 September, 2018.

Description of the VTU-BEC-DB Multimodal Biometrics Database:

Face Database: The face images of VTU-BEC-DB are collected from Sony colour digital camera(i.e. Cyber Shot W610) with a 14.1 megapixel sensor. Face images are collected indoor in a controlled scenarios with white background and variation in illumination and pose. The user's/subjects were seated on a chair with white background while facing a tripod mounted Sony digital camera at a fixed distance. The 4x optical zoom lens of Sony Cyber Shot W610 camera offers a focal length range of 26 - 105mm (35mm equivalent) that helps to capture images from a close distance. The setup for the capturing the face image is made according to the standard interactive distance. The camera was mounted at a height of 120 centimetres while the distance from the camera to the white background was 100 centimetres. Five different orientations i.e. Looking Front (Looking straight into the camera 0^0), Looking Left (around -30^0), Looking Right (around $+30^0$), Looking Up and Looking Down, are considered. There is no restriction for facial expression, so some variances between subject's expressions exist. The pixel resolution of the camera is adjusted to 5 Megapixel and 10 Megapixel separately. Hence total of 30 face images were collected from each subject i.e. 15 facial images for each pixel resolution adjustment. The face images are collected in two sessions in a day by varying the image quality. For every subject, all the 30 facial images were captured on the same day. A total of 3000 facial images were taken from the 100 subjects. The initial resolution of facial images acquired from digital camera were 2592 X 1944 pixels (i.e. for 5 Megapixel adjustment) and 3648 X 2736 pixels (i.e. for 10 Megapixel adjustment). These facial images are resized to 640 X 480 pixels and 800 X 600 pixels respectively.

The facial images are stored in JPEG format with a suitable naming convention as depicted in Figure 1. The first digit of the naming indicates the number of fluorescent lights which are switched in "ON" states i.e. light illumination condition (either two, four and six). Further the next four characters indicate the label of a person i.e. between P001 to P100. The next three characters indicate the pixel resolution of the camera i.e. either M05 or M10 and the last alphabet indicate the sample number i.e. between 1 and 5.
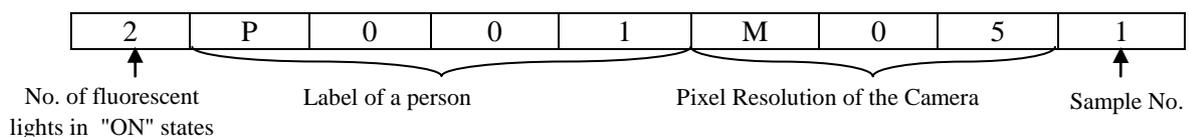
2	P	0	0	1	M	0	5	1

No. of fluorescent lights in "ON" states   Label of a person   Pixel Resolution of the Camera   Sample No.

Figure 1. Naming Convention for The FaceDB

<u>Protocol for Experimentation:</u>

Among the fifteen face images acquired with 5 Megapixel resolution for every user with illumination and pose variations ten face images are selected (i.e. three face images with two lamps ON, namely 2PxxxM051, 2PxxxM053 and 2PxxxM055, four face images with four lamps ON, namely 4PxxxM051, 4PxxxM052, 4PxxxM053 and 4PxxxM054, and three face images with six lamps ON, namely 6PxxxM052, 6PxxxM053 and 6PxxxM054, where 'xxx' indicates the label of a person) and employed for constructing the training set (i.e. 100 persons * 10 samples/ person = 1000 samples) and the remaining five face images are used for test set (i.e. 500 samples). The same experimental protocol is extended for the face images with 10 Megapixel resolution.

Hand Database: The peg free right hand and left hand images are collected from the Book scanner with black-colour background. The scanning speed of the A3 book scanner is set to 0.2 seconds. The cycle time including the image preparation requires 0.7 seconds per hand image with depth focus of 3 inches. The scanning mode is set to 'repetition mode' in which every hand image is stored in JPEG single page file format. From every person 12 samples of the right hand and 12 samples of the left hand images are collected in four sessions over a period of six months. In each session three samples of left hand and three samples of right hand from every person is collected. The hand images are collected in the indoor environment with natural light and also by varying the illumination through fluorescent lights. The light illumination is varied by switching "ON" the one, two and three fluorescent lights separately. For every light illumination variation condition, three hand images are collected. A total of 1200 left hand images and 1200 right hand images were taken from the 100 users/subjects. There are no pegs to restrict the positions of the hand but the fingers are placed apart from each other. The initial resolution of acquired hand images were 4759 X 3307 pixels. These hand images are resized to 1727 X 1200 pixels. The palmprint information can be extracted by cropping the palmprint area from the hand images. The hand images are stored in JPEG format with a naming convention as depicted in Figure 2. The first digit of the name indicates the number of fluorescent lights which are switched in "ON" states i.e. light illumination condition(either zero, one, two and three). The next four characters indicate the label of a person i.e. between P001 to P100. Further the next character indicates either L(left) or R(right) hand. Finally the last alphabet indicate the sample number i.e. between 1 and 3.
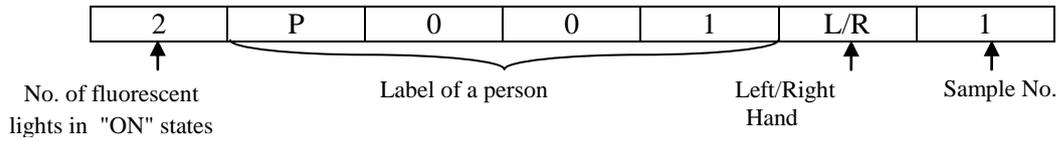
2	P	0	0	1	L/R	1

No. of fluorescent   Label of a person   Left/Right   Sample No.
lights in "ON" states   Hand

Figure 2: Naming convention for the HandDB

Protocol for Experimentation:

The 1200 right hand images are equally partitioned into disjoint sets namely training set and testing set (i.e. each set containing 600 hand images). From every individual six hand images (i.e. two hand images with natural light, namely 0PxxxR1 and 0PxxxR3, one hand image with one lamp ON, namely 1PxxxR2, two hand images with two lamps ON, namely 2PxxxR1 and 2PxxxR3, and one hand image with three lamps ON, namely 3PxxxR2, where 'xxx' indicates the label of a person) are selected for constructing the training set. Further, the remaining six hand images of every individual are used to form the test set. The same experimental protocol is extended for the left hand images.

Voice Corpus: Voice database is collected in two sessions over a period of one year. In each session, five set of recordings of English and five set of recordings of Kannada language number utterance from Zero(0) to Thirty(30) are recorded. A total of 31000 voice samples were collected separately in English and Kannada languages from the 100 users/subjects. The voice samples are collected using Sony ICD- UX533F digital voice recorder. During the recording the voice recorder is adjusted to linear pulse code modulation (LPCM) and the sampling frequency is set to 44.1KHz stereo. The recordings are stored with 16 bit wave files. In order to reduce the ambient noise, the microphone sensitivity is adjusted to medium and the noise cut recording filter (NCF) is employed.

The voice samples are stored in '.wav' format with naming convention as depicted in Figure 3. The first digit of the name indicates the session number i.e. 1 or 2. The next four characters indicate the label of a person i.e. between P001 to P100. The next character indicates either English or Kannada language and a digit indicates the recordings viz. in each session, five set of recordings of English and five set of recordings of Kannada language numbers. The remaining two digits provide the sample number of voiceprint between Zero(00) to Thirty(30).
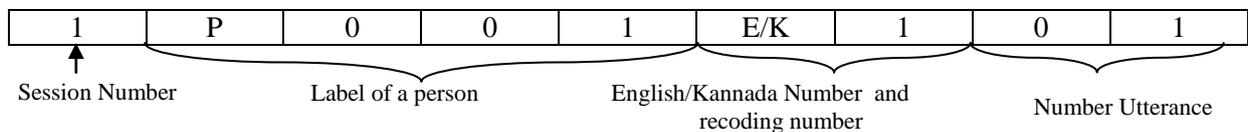
1	P	0	0	1	E/K	1	0	1

Session Number   Label of a person   English/Kannada Number and   Number Utterance
                                            recoding number

Figure 3: Naming convention for the VoiceDB

Protocol for Experimentation:

The voice corpus contains 62000 number utterances (i.e. 31000 voiceprints separately in English and Kannada languages) are recorded from 100 subjects in two sessions over a period of one year. In each session, five set of recordings of English and five set of recordings of Kannada language number utterance from Zero(0) to Thirty(30) are recorded. For English voice corpus, from each speaker five utterances (i.e. three voiceprints of first session namely 1PxxxE1yy, 1PxxxE3yy and 1PxxxE5yy, and two voiceprints of second session namely 2PxxxE7yy, 2PxxxE9yy where 'xxx' indicates the label of a person and 'yy' indicates the number uttered) are used for constructing the training set(5 voiceprints/person * 100 person = 500 voiceprints of each number utterance) and the remaining five utterances are employed for testing (500 voiceprints of each number utterance).

Similarly for Kannada voice corpus, from each speaker five utterances (i.e. three voiceprints of first session namely 1PxxxK1yy, 1PxxxK3yy and 1PxxxK5yy, and two voiceprints of second session namely 2PxxxK7yy, 2PxxxK9yy where 'xxx' indicates the label of a person and 'yy' indicates the number uttered) are used for constructing the training set(5 voiceprints/person * 100 person = 500 voiceprints of each number utterance) and the remaining five utterances are employed for testing (500 voiceprints of each number utterance). In the similar manner the protocol is extended for the remaining utterances of English and Kannada language number utterances.

Description of the File Storage:

The collected multimodal biometric database is organized in the tree like directory structure and is described in the following Table 2. The database is stored in the tree structure format with a *'VTU-BEC-DB'* as a parent root directory. The biometric data collected from the face, hand and voice traits are stored in three root directories namely *'FaceDB'*, *'HandDB'* and *'VoiceDB'* respectively and are the part of parent directory. Further the face images acquired with adjusting camera resolution to 5 Megapixel and 10 Megapixel by variations in pose and illumination are stored in *'5M'* and *'10M'* sub-directories respectively in the *'FaceDB'* directory.

Table 2: Description of the File Storage - Sample Database

Name of Parent Root Directory	Name of Root Directory	Name of Sub-Directory	Name of Sub-Directory	Typical Name of the File
VTU-BEC-DB	FaceDB	5M	-	2P001M051.jpg
				4P046M052.jpg
				6P100M053.jpg
		10M	-	2P007M101.jpg
				4P033M103.jpg
				6P095M105.jpg
	HandDB	L	-	0P001L1.jpg
				1P016L2.jpg
				2P062L3.jpg
				3P100L2.jpg
		R	-	0P001R1.jpg
				1P025R2.jpg
				2P068R3.jpg
				3P100R2.jpg
	VoiceDB	English	S1	1P001E126.WAV
				1P099E509.WAV
			S2	2P001E600.WAV
				2P073E630.WAV
		Kannada	S1	1P001K121.WAV
				1P095K527.WAV
			S2	2P011K223.WAV
				2P011K423.WAV

The hand images collected from the left and right hand with a variations in illumination are stored in *'L'* and *'R'* sub-directories respectively in the *'HandDB'* directory. The palmprint images can be easily collected from the hand images. Further the voiceprints uttered in English and Kannada languages are stored in the *'English'* and *'Kannada'* sub-directories respectively in the *'VoiceDB'* directory. The voiceprints in each of the language is recorded in two sessions and stored in *'S1'* and *'S2'* sub-directories in their respective directories viz. *'English'* and *'Kannada'* sub-directories. The files are stored in the leaf directories according to the naming conventions.

For further information contact:

(1) Dr. Shanmukhappa A. Angadi
    email-id : saangadi@vtu.ac.in, vinay_angadi@yahoo.com

(2) Mr. Sanjeevakumar M. Hatture
    email-id : smhatture@yahoo.com, smhatture@gmail.com