Article

3DAirSig: A Framework for Enabling In-Air Signatures Using a Multi-Modal Depth Sensor

Jameel Malik [1,2,3], Ahmed Elhayek [4], Sheraz Ahmed [1,*], Faisal Shafait [3,5], Muhammad Imran Malik [3,5,*] and Didier Stricker [1,2]

[1]   German Research Center for Artificial Intelligence, DFKI, Kaiserslautern 67653, Germany; jameel.malik@dfki.de (J.M.); Didier.Stricker@dfki.de (D.S.)

[2]   Department of Informatics, University of Kaiserslautern, Kaiserslautern 67653, Germany

[3]   School of Electrical Engineering and Computer Science (SEECS), National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan; faisal.shafait@seecs.edu.pk

[4]   University of Prince Mugrin (UPM), Madinah 20012, Saudi Arabia; ahmed.elhayek@dfki.de

[5]   Deep Learning Laboratory, National Center of Artificial Intelligence (NCAI), Islamabad 44000, Pakistan

*   Correspondence: sheraz.ahmed@dfki.de (S.A.); malik.imran@seecs.edu.pk (M.I.M)

Abstract: In-air signature is a new modality which is essential for user authentication and access control in noncontact mode and has been actively studied in recent years. However, it has been treated as a conventional online signature, which is essentially a 2D spatial representation. Notably, this modality bears a lot more potential due to an important hidden depth feature. Existing methods for in-air signature verification neither capture this unique depth feature explicitly nor fully explore its potential in verification. Moreover, these methods are based on heuristic approaches for fingertip or hand palm center detection, which are not feasible in practice. Inspired by the great progress in deep-learning-based hand pose estimation, we propose a real-time in-air signature acquisition method which estimates hand joint positions in 3D using a single depth image. The predicted 3D position of fingertip is recorded for each frame. We present four different implementations of a verification module, which are based on the extracted depth and spatial features. An ablation study was performed to explore the impact of the depth feature in particular. For matching, we employed the most commonly used multidimensional dynamic time warping (MD-DTW) algorithm. We created a new database which contains 600 signatures recorded from 15 different subjects. Extensive evaluations were performed on our database. Our method, called 3DAirSig, achieved an equal error rate (EER) of 0.46%. Experiments showed that depth itself is an important feature, which is sufficient for in-air signature verification. The dataset will be publicly available (https://goo.gl/yFdfdL).

Keywords: in-air signature; depth sensor; convolutional neural network (CNN); 3D hand pose estimation; multidimensional dynamic time warping (MD-DTW)

1. Introduction

Electronic identity authentication plays a vital role for access control and security in modern age. In e-authentication, a protected token (e.g., a cryptographic key) is used to access a system or an application on a network. Biometric-based authentication uses physical, behavioral, or adhered human characteristics for identification. These characteristics include, for instance, a fingerprint, iris scan, handwritten signature, color, gait, and facial scan. Biometric authentication is more secure and less prone to identity theft [1]. With the rapid growth of technology, emerging concepts, such as *classroom of the future* http://iql-lab.de [2], would allow smart interactions in a virtual and augmented reality environment. In such a noncontact mode of interaction, biometric in-air signature verification is important for access control and authentication. Traditionally, signature verification methods

are classified into two types, namely, offline and online signature verification. In offline signature verification, a handwritten signature is acquired on a document and verified using a scanned or camera-captured image of the 2D signature [3–5]. The artificial neural network (ANN), support vector machine (SVM), and pixel matching technique (PMT) are famous classification algorithms, which have been used by offline methods. On the other hand, in online methods, e-signatures are taken on a touch device (e.g., tablet or pad) using an e-pen or finger movement on a digital screen [6–13]. These methods are difficult to forge due to various dynamic features, such as velocity, acceleration, and pen pressure. The signature acquisition techniques mentioned above exploit the 2D spatial and temporal information taken from a digital touch screen or a document. For verification, dynamic time warping (DTW) is the most effective and widely used technique [14,15], mainly because of its ability to well align temporal signals. Other prominent approaches based on a. neural network (NN) [12], SVM [13], and the hidden Markov model (HMM) [9] have also been employed for online verification.

In-air signatures are a new modality which allows a user to sign in the air by making free hand movements, thereby eliminating the need for a writing surface. Notably, this modality inherently contains important information in the third dimension (i.e., depth), in addition to the 2D spatial pattern. Existing methods for in-air signature verification use either an RGB or depth camera, a wearable camera (e.g., Google Glass) or a movement sensor in a cell phone [1,16–18]. However, these methods address the problem of in-air signature acquisition and verification in the conventional way. More precisely, the focus of these approaches has been inclined towards the utilization of the 2D spatial and temporal features. Lack of consideration towards the hidden depth information has restricted the exploration of the full potential in the 3D signature trajectory. In this work, we investigate the potential of the unique depth pattern. We show that the depth itself is a strong feature, which is sufficient for in-air signature verification. On the other hand, fingertip tracking is a challenging problem, especially due to the occlusions of fingers and viewpoint changes during signing freely in the air. The acquisition of a correct in-air signature trajectory is crucial to verification. This problem has not been well-addressed because the existing approaches try to locate only the fingertip using heuristics. Some of the approaches rely on palm center point tracking [17,19] which does not accurately mimic the pointing finger movement while signing in the air. Furthermore, due to their complex in-air signature acquisition systems, they are not suitable for real-time applications. In principle, the skeleton of a human hand is a kinematic structure where each child joint is connected to its parent joints [20,21]. Therefore, for a stable and reliable tracking of the position of a fingertip, the complete 3D pose of a hand should be estimated. In contrast to existing fingertip-tracking approaches, we exploited the huge progress of the convolutional-neural-network (CNN) based hand pose estimation using a low cost multimodal depth sensor [22] and trained a CNN to estimate the hand joints' keypoints in 3D; see Section 4.3. Estimating a full hand pose is more stable, especially in the case of occluded fingertips, as it learns to estimate all features of the hand. We created our own database of in-air signatures for analysis and verification. The dataset will be publicly available. We performed a detailed ablation study, which especially reveals the significance of the hidden depth feature in verification. We propose an improved spatial-features-based verification strategy which incorporates the depth information; see Section 6.1. We employed the most common and effective multidimensional dynamic time warping (MD-DTW) algorithm for matching, since our focus is to investigate and highlight the potential in individual features of the in-air signature using the best practice for verification.

2. Related Work

Comprehensive reviews on offline and online signature verification have been reported in References [23–25]. Keeping in view the relevance with our work, here we discuss the published literature on in-air signature verification. Katagiri et al. [26] proposed the first free space personal authentication system. They adopted a high-speed video camera to acquire an in-air signature trajectory. For verification, they employed a commercial signature verification engine provided by CyberSIGN Japan Inc. (Tokyo, Japan) http://www.cybersign.com. In Reference [27], Takeuchi et al. combined

hand shape features with an RGB camera to capture handwriting motion in the air. Keeping in view the extended use of smartphones in various applications, Diep et al. [28] used a motion sensor in a smartphone to record signature data. They used SVM for verification. Matsuo et al. [29] introduced an adaptive template update method in order to improve long-term stability in arm-swing motion. Jeon et al. [17] adapted a low-cost depth camera to capture an in-air signature trajectory. In order to record the signature trajectory, they introduced a heuristic approach to detect the palm center position. Bailador et al. [18] investigated various pattern recognition techniques, i.e., HMM, Bayes classifier, and DTW, for authentication. The best performance was shown by the DTW algorithm. In order to capture in-air signature trajectory, the authors used an embedded 3D accelerometer in a mobile phone. With the recent trend towards wearable technology, Sajid et al. [1] proposed a new in-air signature acquisition method using Google Glass. They used a motion-based video segmentation algorithm along with a skin-color-based hand segmentation in order to acquire signature data. A video-based in-air signature verification system using a high-speed RGB camera was introduced by Fang et al. [16]. They traced the fingertip using an improved tracking learning detection (TLD) algorithm. For the verification phase, the authors developed a fusion algorithm based on an improved DTW and the fast Fourier transform (FFT). Recently, Khoh et al. [19] proposed a predictive palm segmentation algorithm to create a motion history image (MHI) using a depth sensor. Afterwards, they produced a two-dimensional representation of a hand-gesture signature based on the MHI. All of the methods mentioned above treat and process in-air signature trajectories in the conventional online form. However, we emphasize that in-air signatures enclose a unique hidden depth feature, which should not be ignored in acquisition and verification. In this work, we investigate the potential of this important feature. On the other hand, the reported methods for fingertip tracking are based on heuristics, which are not feasible for practical applications. Inspired by the recent progress in deep-learning-based hand pose estimation using a depth sensor [22], we propose a new real-time algorithm for in-air acquisition which regresses the 3D hand pose rather than detecting only the fingertip or palm center. Therefore, the proposed method is not restricted to any specific hand pose and has the ability to perform well in cases of occlusion.

3. Framework Overview

The block diagram of our proposed 3D in-air signature acquisition and verification framework is shown in Figure 1. For the signature acquisition, we propose a CNN-based hand pose estimation method to predict the 3D hand joint positions from a single depth image. The input depth frame D_i is captured using Intel's creative senz3D depth camera [30]; see Section 4.1 for details of our acquisition setup. The hand region is segmented from D_i using center of hand mass (CoM) followed by a crop function; see Section 4.2. The output D_s is fed to the PoseCNN, which predicts the 3D hand pose; see Section 4.3. The estimated joint position of the index fingertip in each depth frame is used to record the 3D signature trajectory. The recorded in-air signature trajectory is preprocessed for normalization and smoothing; see Section 5.1. Thereafter, spatial and depth features are extracted from the 3D signature. For matching, MD-DTW is used to obtain a similarity measure between the selected feature of the preprocessed test signature and the corresponding precomputed feature template. In the final step, the test signature is verified by the decision threshold; see Sections 5.3 and 5.4.

4. In-Air Signature Acquisition

In this section, we explain our 3D in-air signature acquisition setup, fingertip-tracking approach, and the dataset creation.

4.1. Data Acquisition Setup

Figure 2 shows our in-air signature acquisition setup. A user is allowed to sign freely in the air within the field of view (FoV) of Intel's creative senz3D depth camera mounted on top of the screen. The FoV of the camera is 74° diagonal. Two position markers are placed on either side of the depth camera to provide an approximate start and end position for recording the signature.

Our acquisition system allows to easily select between left or right hand before signing. During the signature acquisition, the user's hand should be the closest object to the camera. Notably, our method is not restricted to a specific hand pose for signing in the air. However, most of the users participating in our database creation used a natural pointing index finger pose (as shown in Figure 1). Our system allows a user to see a 2D projection of the 3D signature trajectory in real-time on a signature pad, which is displayed on a monitor screen. Our acquisition system is robust to variations in ambient light intensity in indoor environments.
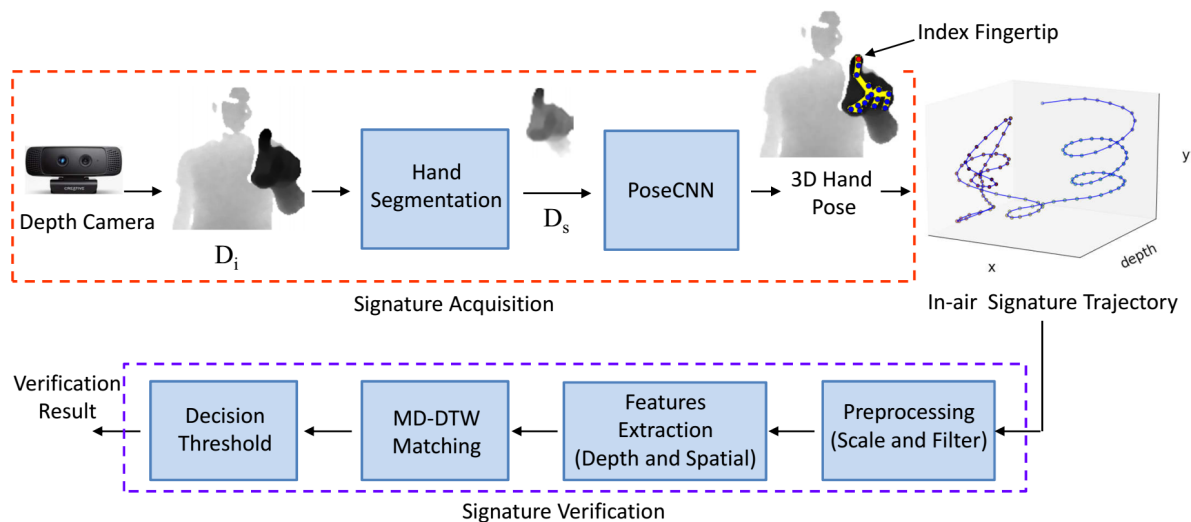

Figure 1. An overview of our method for in-air signature acquisition and verification. In the acquisition phase, the hand region is first segmented from a raw depth frame. Then, the estimated 3D position of the index fingertip is recorded for every frame using a CNN-based hand pose estimation method. For verification, the test signature is scaled and filtered. Thereafter, the spatial and depth features are extracted for matching using the MD-DTW algorithm. Finally, the test signature is verified by the decision threshold.

Figure 2. Our setup for in-air signature acquisition. The depth camera is mounted on top of the screen. The position markers on both sides of the depth camera allow capturing of in-air signature within the field of view (FoV) of the camera. Three GoPro cameras are placed around a user to record the hand motion in 3D space from different view points. Camera 3 specifically records the depth variation.

4.2. Hand Segmentation

An accurate segmentation of the hand region from a raw depth frame is important for learning-based hand pose estimation approaches. We used a hand segmentation method similar to that described in Reference [31] (Figure 3a). The segmentation process has two steps. The first step is to find an accurate 3D location of the hand palm center. As mentioned earlier, the hand is assumed to be the closest object to the camera; therefore, a simple depth value-based thresholding

can be used to separate the human body from the hand. We used a depth threshold of 600 mm. Then, the 3D location of the palm center is calculated by averaging all the pixels which belong to the hand region (i.e., pixel values less than 600 mm). The second step is to preprocess or crop the hand region in 3D using the obtained palm center. In Figure 3a, the function f crops the hand region around the calculated palm center using a bounding box. The size of the bounding box is 150 mm. Then, depth values are normalized to $[-1, 1]$. The resultant image is of a size of 96×96. The runtime of our hand segmentation method is 0.47 ms.

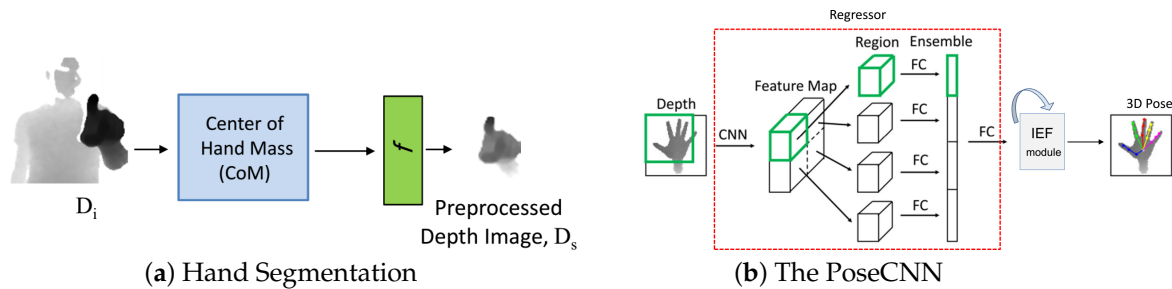(**a**) Hand Segmentation　　　　　　　　　　(**b**) The PoseCNN

Figure 3. (**a**) shows our approach for hand segmentation from a raw depth frame. First, the center of hand mass (CoM) is calculated, provided that the hand is the closest object to the depth camera. Then, the function f crops the hand region in 3D. (**b**) The PoseCNN takes the cropped hand image as input and regresses 3D sparse joints keypoints.

4.3. Fingertip Tracking

Stable and reliable fingertip tracking is essential for the correct recording of a 3D in-air signature. For this purpose, we exploited the huge progress of CNN-based hand pose estimation methods. One of the major advantages associated with these methods is that they estimate the complete hand pose rather than detecting only the fingertip or palm center. This is particularly important in cases of severe occlusions of fingers during signing in the air. An overview of our method is shown in Figure 3b. The PoseCNN is used to estimate the 16 3D joint positions of the hand skeleton from a single depth image. The first part of the PoseCNN (i.e., Regressor) is adopted from [31], which originally regressed 3D hand poses using a single shared CNN for feature extraction and a powerful yet simple region ensemble (REN) strategy. In our implementation, the final fully connected (FC) layer of the regressor outputs features $\varphi \in \mathbb{R}^{512}$ instead of joint positions.

Architecture of the Regressor: The architecture of the shared CNN for feature extraction comprises six convolution layers using 3×3 kernel sizes. A rectified Linear Unit (ReLu) is connected with each of the convolution layers as an activation function. A max pooling layer with a stride of 2 is connected after every consecutive pair of convolution layers. Two residual connections are incorporated between the pooling layers. The output features are of size $12 \times 12 \times 64$. Then, two FC layers of dimension 2048 are connected with a dropout ratio of 0.5. As shown in Figure 3b, the feature maps from different regions of the input depth image are divided into a 2×2 grid. Thereafter, the features from the FC layers of the grid regions are simply concatenated. The final FC layer after the concatenation produces $\varphi \in \mathbb{R}^{512}$. We refer the reader to Reference [31] for further details of the shared CNN architecture and the REN strategy.

IEF module: We integrate an iterative error feedback (IEF) module to the end of the regressor for refinement of the estimated hand pose. The output of the regressor φ is concatenated with an initial estimate of hand pose H_p i.e., $\phi = \{\varphi, H_p\}$. H_p is obtained by averaging all the joint positions from the ground truth annotations of the datasets. ϕ is fed to the IEF module, which comprises two FC layers with 512 neurons each. Both the FC layers use dropout layers with a ratio of 0.3. The last FC layer contains 48 neurons, corresponding to the 16 3D joint positions. The IEF module basically refines H_p in an iterative feedback manner such that $H_p(t+1) = H_p(t) + \delta H_p(t)$. We use three iterations.

Training of the PoseCNN: In order to improve the generic performance of the PoseCNN, especially for varying hand shapes, we trained on a combined dataset (i.e., *HandSet*) proposed in Reference [21]. The *HandSet* encapsulates three famous public hand pose datasets in a single unified format. These datasets include NYU [32], ICVL [33], and MSRA-2015 [34]. Our network runs on a desktop using Nvidia's Geforce GTX 1080 Ti GPU. We used a learning rate (LR) of 0.001 with a 0.9 stochastic gradient descent (SGD) momentum and a batch size of 256. One forward pass through the PoseCNN takes 3.2 ms.

Accuracy of predicted fingertips positions: We quantitatively evaluated the accuracy of estimated fingertips positions on the NYU test dataset. The 3D joint location error on fingertips comes out to be 13.2 mm, which is better than the lowest reported error (15.6 mm) in Reference [35]; see Table 1.

Table 1. The table shows the mean 3D joint location error (mm) for fingertips of various methods on the NYU [32] hand pose test dataset.

Method	Fingertips 3D Joint Location Error
DeepModel [20]	24.4 mm
Oberweger et al. [36]	23.2 mm
REN [35]	15.6 mm
Ours	13.2 mm

4.4. The Dataset Creation

There are two main motivations for creating our dataset for in-air signature verification. The first is to study the potential of the hidden depth feature. The second is to exploit the great progress in CNN-based hand pose estimation for stable and reliable fingertip tracking. For video recordings of genuine signatures which are shown to impostors, we used three GoPro cameras in our capture setup; see Figure 2. Two of the cameras (Cameras 1 and 2) were placed behind and right-front of the subject to record the spatial pattern of the signature. The third camera (Camera 3) recorded from the side view to visualize the depth variation in the signature. The users were asked to practice multiple times before the actual recordings as signing in the air is generally not a well-familiar modality. We emphasized on making explicit variations in depth during signing, which allows to fully exploit the hidden depth feature in the in-air signature trajectory. Our database (the dataset will be publicly available at https://goo.gl/yFdfdL) includes 600 signatures from 15 users. We recorded 15 genuine signatures from each of the users and obtained 25 forgeries for every original writer from 5 impostors. Ten out of 15 genuine signatures were used for the testing phase and the remaining were used for the training phase; see Section 5. Samples of genuine preprocessed signatures with the corresponding 2D spatial views and unique depth patterns are shown in Figure 4. The color variations in the 3D view of a signature show variation in the depth pattern; see Figure 4a. Notably, each signature has a unique depth pattern (Figure 4c) which is challenging to forge jointly with the spatial pattern; see Section 6.
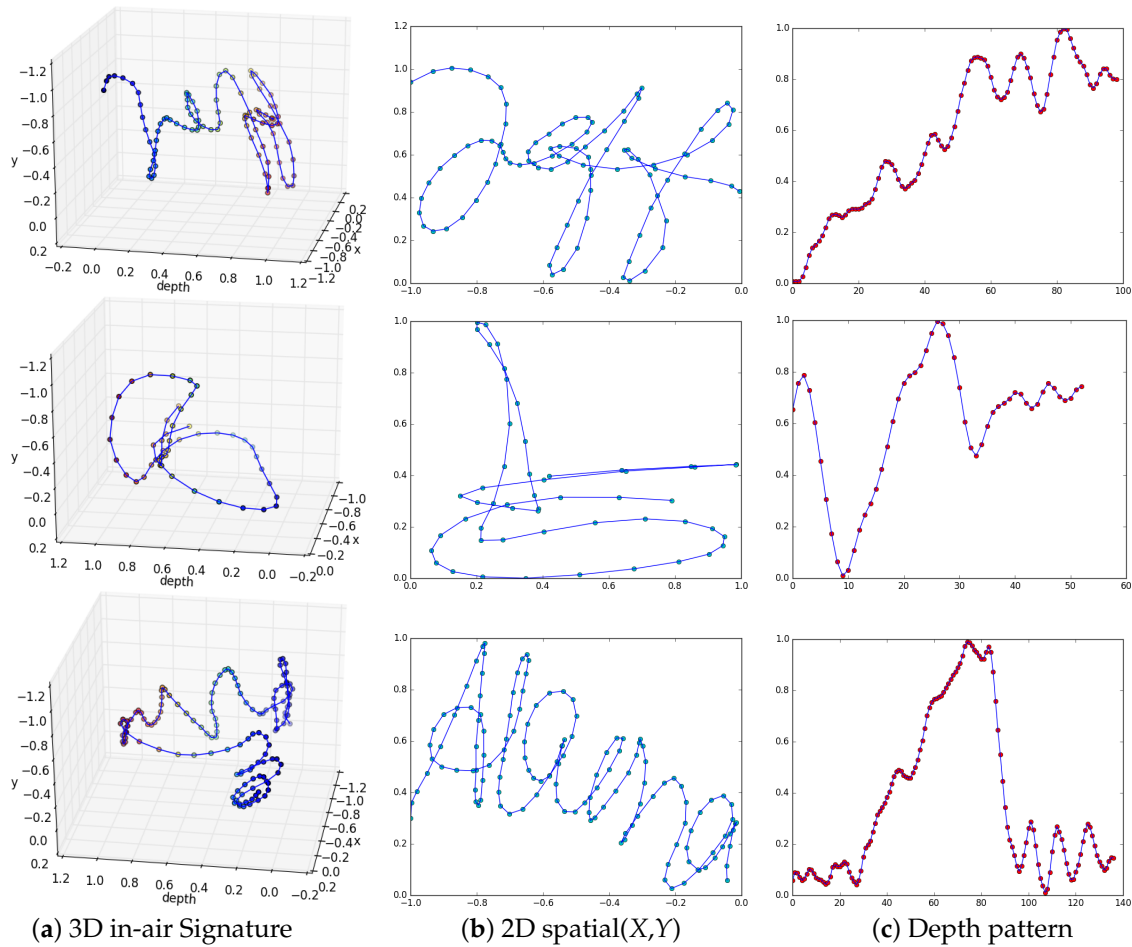
(**a**) 3D in-air Signature   (**b**) 2D spatial(*X*,*Y*)   (**c**) Depth pattern

Figure 4. Samples of genuine in-air signatures from our dataset. Each one of the rows shows (**a**) the 3D in-air signature trajectory, (**b**) the 2D spatial view, and (**c**) depth pattern. The depth pattern of each signature is particularly unique and, therefore, it is an important hidden feature.

5. In-air Signature Verification

In this section, we explain the preprocessing, extracted features, training, and testing phases. We adopted a commonly used MD-DTW algorithm for matching, mainly because it can align temporal signals well even though they are not consistent in time.

5.1. Preprocessing

The recorded in-air signature is preprocessed for normalization and smoothing. An appropriate preprocessing of a signature can affect the results of signature verification [11,17]. First, we removed a few redundant 3D points from the start and end of a signature trajectory whose displacement was less than 3 pixels. The removed points corresponded to a small wait time before starting the actual hand motion and a time to close the recording after the end of the signature. In order to remove discontinuities due to fast hand movements, we applied a moving average filter with a window size of 5, which resulted in a smoother signature trajectory. Thereafter, we normalized the signatures to compensate for variations in *position* and *scale*. For normalization, the transformation from absolute to relative values in 3D can be obtained using the following formulas:

$$X_j^* = (X_j - X_{min})/(X_{max} - X_{min}) \tag{1}$$

$$Y_j^* = (Y_j - Y_{min})/(Y_{max} - Y_{min}) \tag{2}$$

$$Z_j^* = (Z_j - Z_{min})/(Z_{max} - Z_{min}), \quad\quad (3)$$

where X_j, Y_j, and Z_j are the original or absolute values of a signature. X_j^*, Y_j^*, and Z_j^* are the transformed values. $X_{min}, X_{max}, Y_{min}, Y_{max}, Z_{min}$, and Z_{max} are the minimum and maximum values of X_j, Y_j, and Z_j. A test signature before and after the preprocessing step is shown in Figure 5.
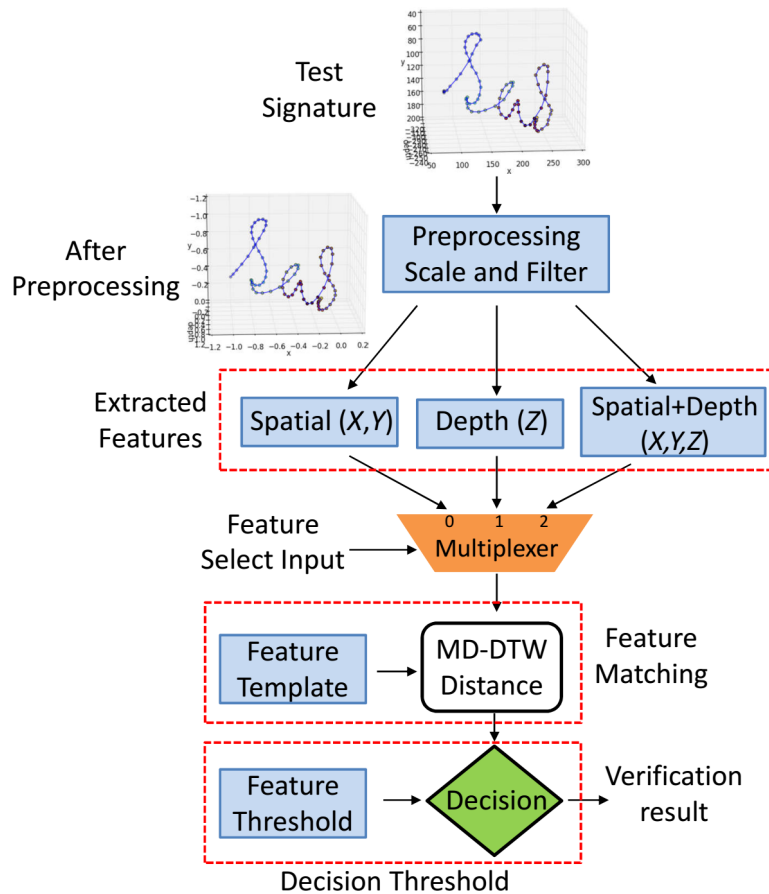

Figure 5. The flow diagram of the testing phase of our in-air signature verification system. The test signature is preprocessed for normalization and smoothing. The extracted features include spatial, depth, and spatial plus depth. Then, a multiplexer with a control input is used to select one of the extracted features. The selected feature is matched with the corresponding feature template using the MD-DTW algorithm. Finally, the verification result is produced by the decision threshold.

5.2. Feature Extraction

Figure 6 shows all the feature combinations we used in our verification process. We studied the impact of the hidden depth feature in different ways. The spatial (X,Y) is a commonly used 2D representation of in-air signatures; see Figure 6b. However, we argue that only the spatial (X,Y) is not a complete representative of an in-air signature trajectory. Therefore, we extracted two new types of spatial features, i.e., spatial (X,Z) and spatial (Y,Z) which implicitly incorporate the depth feature. We also studied the impact of these two features when combined with the spatial (X,Y); see Section 6. Nevertheless, the most interesting feature is the hidden depth pattern (Figure 6e) which has not been fully explored in the previous works.
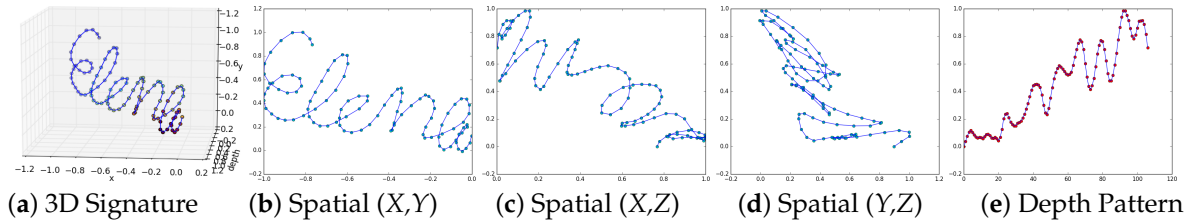
(**a**) 3D Signature　　(**b**) Spatial (*X,Y*)　　(**c**) Spatial (*X,Z*)　　(**d**) Spatial (*Y,Z*)　　(**e**) Depth Pattern

Figure 6. Illustration of different features which are used for in-air signature verification. We fully exploited different combinations of the features inherently present in the in-air signature trajectory to improve the performance of the verification system. The unique depth feature of a user especially plays a vital role in verification phase.

5.3. Training Phase

In this phase, we computed the feature templates and the respective feature thresholds using 75 genuine training samples. We used neither forgeries nor original signatures from the test set. It is worth noting that many pattern recognition researchers use models, e.g., NN, SVM, while training them on the positive (genuine) and negative (forgery) samples at the same time [37,38]. According to forensic handwriting examiners [39], this is unrealistic as, in the real world, one can never limit the forgery set and every signature, other than the concerned genuine signatures, can be considered a forgery. Furthermore, in real forensic cases, a verification system can only have genuine specimen samples and one or more questioned signatures. Henceforth, the best approach while using such models is to train them only on genuine specimen signatures. This can be done using specialized one class classifiers, like SVM/NN, for one class classification [40–43]. As explained earlier, we used five features; see Figure 6. Hence, a total of five feature templates and five respective feature thresholds for each of the 15 users are computed. A feature template is generated by averaging the features of the five training samples. We calculated a feature threshold value from five training samples of a signee, which are reserved for the training phase using the 4-fold cross validation strategy (i.e., using limited signatures for estimating how the system will perform when used to make predictions on data not used during training.

4-fold cross validation strategy: In this methodology, we randomly shuffled five genuine training signature samples and divided them into two groups. The first group contained four training samples, which were taken as the training set. The second group contained only one training sample, which was considered the dummy test set. More specifically, let $\mathbf{S} = \{S_{t_1}, S_{t_2}, S_{t_3}, S_{t_4}, S_{t_5}\}$ be the five training samples of a signature, where $S_x \in R^{d \times L_x}$. L_x is the length of the signal S_x and d is the number of dimensions of one point in the signal. In the first round, we split $\mathbf{S}$ into two subsets, $\mathbf{S}_a = \{S_{t_2}, S_{t_3}, S_{t_4}, S_{t_5}\}$ and $\mathbf{S}_b = \{S_{t_1}\}$. This is simply taking the first sample S_{t_1} out of comparison in this round. For $\mathbf{S}_a$, we make a 4 x 4 confusion matrix $\mathbf{C}_1$ using Equations (4) and (5). From $\mathbf{C}_1$, we manually select a threshold value th_1 such that any compared threshold value greater than th_1 will declare the signature as forged. In the second round, we eliminate S_{t_2} and calculate another 4 x 4 matrix $\mathbf{C}_2$ and find th_2 . In a similar way, we calculate $\mathbf{C}_3$, $\mathbf{C}_4$, and $\mathbf{C}_5$ and select the respective thresholds th_3, th_4, and th_5. Finally, we simply take the mean th_m of these five threshold values. The th_m is used in the final decision threshold process.

5.4. Testing Phase

Figure 5 shows the flow chart of the testing phase. After the preprocessing step and the feature extraction, a feature select input of a 3 x 1 multiplexer allows to select one of the features, i.e., spatial, depth, or spatial plus depth. After the selection of a desired feature, a similarity measure is found with the corresponding feature template using the MD-DTW algorithm [44] as follows:

MD-DTW Matching: Let $s_1 \in R^{d \times L_{s_1}}$ and $s_2 \in R^{d \times L_{s_2}}$ be the two time series signals, where L_{s_1} and L_{s_2} are the lengths of s_1 and s_2, respectively, and d is the dimension of a single point in the signal. The distance matrix M(i,j) can be computed using the L2-norm without square root operation as:

$$M(i,j) = \sum_{k=1}^{d} (s_1(k,i) - s_2(k,j))^2. \tag{4}$$

After obtaining the matrix M(i,j), the distance or similarity score between the elements of s_1 and s_2 on the DTW path can be found using the following equation:

$$D(i,j) = M(i,j) + \min \begin{cases} D(i\text{-}1,j) \\ D(i\text{-}1,j\text{-}1) \\ D(i,j\text{-}1) \end{cases} \tag{5}$$

Decision Threshold: In the final step, as shown in Figure 5, the obtained similarity score is simply compared with the corresponding feature threshold th_m; see Section 5.3. The test signature is verified if the DTW distance is less than the feature threshold.

6. Experiments and Results

In this section, we detail the experiments performed on our dataset. The performances are reported using the false rejection rate (FRR), false acceptance rate (FAR), and equal error rate (EER) as evaluation metrics.

6.1. Ablation Study

In this subsection, we detail the ablation study, which was performed on the extracted features (Figure 6). The impact of every feature on the performance of verification was investigated and the results are reported on our captured dataset. We propose four different implementations of a verification module based on the extracted features from the in-air signature trajectory. *Depth-based signature verification (DSV) module*: To study the effectiveness of the hidden depth feature in verification, we implemented the verification module based on only the 1D depth Z of the signature trajectory. In Figure 5, the feature select input of the multiplexer is set to 1 in order to select the extracted depth feature from the test signature. The distance measure between the depth feature of the test signature and the precomputed depth feature template was calculated using Equations (4) and (5). The obtained similarity score was compared with the precomputed depth feature threshold to verify the test signature. Quantitative results on individual users are shown in Table 2. In Table 3, the *DSV* module shows FAR, FRR, and EER of 1.33%, 2.00%, and 0.51%, respectively. Qualitatively, the depth patterns of the genuine and forged signatures are shown in Figure 7. Despite the fact that the spatial patterns of the forgeries are closer to the genuine signatures, the depth patterns are distinct. As mentioned ealier, the impostors were shown the video recordings of the signatures from different camera views. However, they were either unable to notice exact variations in depth or it was difficult to forge the depth pattern. These results show the importance of the depth feature, which alone can provide a reliable verification. We also observed that it is more challenging for the impostor to forge the depth pattern simultaneously with the spatial pattern.
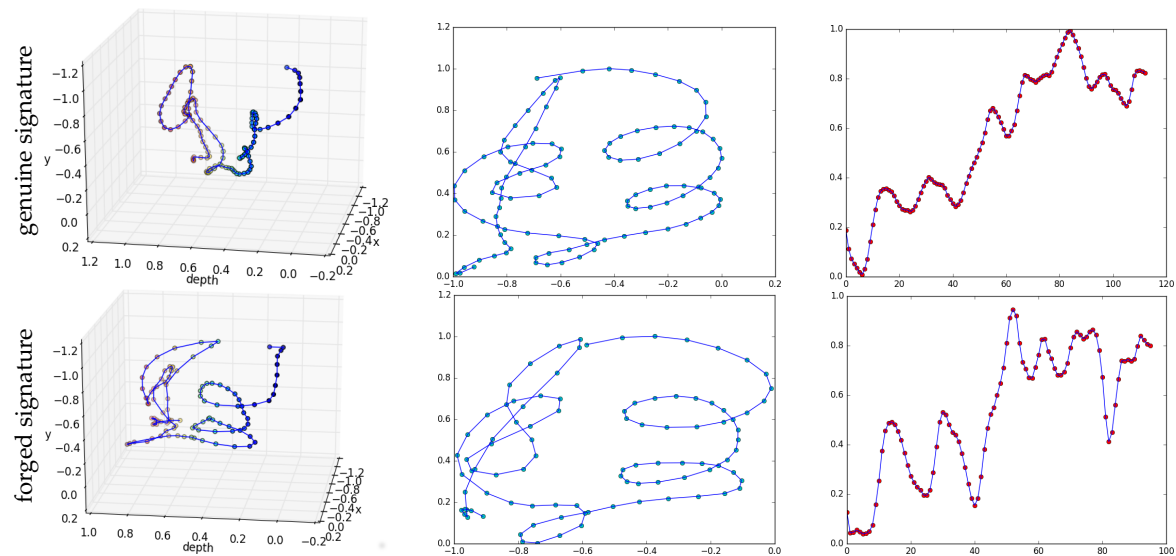
Figure 7. Comparison of spatial and depth patterns of the genuine and the corresponding forged signature. The top row shows a sample of a genuine signature and its corresponding spatial and depth patterns and the bottom one shows the respective forged signature. The color change shows the variation in depth pattern (3D view in the first column). Clearly, the depth pattern of the forged signature is different than the original one, although spatially they seem to be close.

2D spatial-based signature verification (SSV) module: We implemented this verification module using only the 2D spatial (*X,Y*) feature; see Figure 6b. The feature select input of the multiplexer was set to 0; see Figure 5. The similarity score between the extracted spatial feature of the test signature and the spatial (*X,Y*) feature template was obtained using Equations (4) and (5). Then, the DTW distance was compared to the spatial feature threshold for the verification. Quantitative results are shown in Tables 2 and 3. The performance of this verification module shows that considering only the spatial feature (*X,Y*) of the in-air signature trajectory results in a larger number of false acceptances and false rejections, thereby producing higher error rates.

Table 2. The table shows the results of the four verification modules on our dataset. The number of false rejections (FR), false acceptances (FA), and total errors are provided for each of the 15 users. The *3D-SV* module shows the least number of FA, while its number of FR is equivalent to the *DSV* module.

Subject	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Error Number
DSV																
FR	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	3
FA	0	1	0	0	0	0	0	0	0	0	1	0	1	2	0	5
SSV																
FR	0	1	0	0	0	0	0	1	0	0	0	1	2	0	3	8
FA	0	2	0	0	0	1	0	2	0	0	2	0	1	3	0	11
ISSV																
FR	0	1	0	0	0	0	0	0	0	0	0	1	1	0	2	5
FA	0	1	0	0	0	0	0	1	0	0	1	0	1	2	0	6
3D-SV																
FR	0	1	0	0	0	0	0	0	0	0	0	1	0	0	1	3
FA	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	3

Table 3. The table shows the person independent FAR, FRR, and EER for each of the four verification modules. There are a total of 150 genuine test and 375 forged signatures. The *3D-SV* module shows the best results, while the *DSV* module demonstrates competitive performance.

Verification Module	FAR (%)	FRR (%)	EER (%)
DSV	1.33	2.00	0.51
SSV	2.93	5.33	0.69
ISSV	1.60	3.34	0.58
3D-SV	0.80	2.00	0.46

Improved 2D spatial-based signature verification (ISSV) module: We attempted to improve the performance of the *SSV* module by incorporating additional spatial feature combinations (i.e., Spatial (*X,Z*) and Spatial (*Y,Z*)). The block diagram of the *ISSV* module is shown in Figure 8. The DTW matching is performed on these additional features in parallel to the traditional spatial (*X,Y*) using precomputed respective feature templates. Thereafter, binary decisions were obtained for each individual feature using the corresponding feature thresholds. Lastly, the final verification result was produced by a simple majority voting scheme, which declared the test signature as verified if no less than 2 features passed the corresponding decision thresholds. The verification results are reported in Tables 2 and 3 that clearly show an improved performance compared to the *SSV* module. There is a notable reduction in the number of false acceptances and false rejections. The EER is reduced by 15.9% compared to the *SSV* module. However, the performance is still lagging behind the *DSV* module.
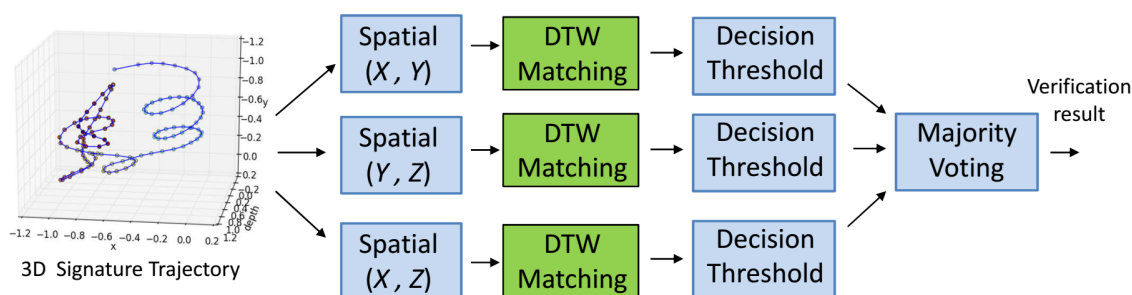

Figure 8. Our framework for an improved 2D spatial-based signature verification (*ISSV*) module. The spatial features (i.e., *(X,Y)*, *(Y,Z)* and *(X,Z)*) are separately matched with the respective precomputed feature templates using a 2D-DTW algorithm. Thereafter, binary decisions are made by the decision thresholds. Lastly, the test signature is verified using a simple majority voting scheme.

3D signature verification (3D-SV) module: In this verification module, we exploited the full 3D information (i.e., *X, Y, Z*) altogether. In Figure 5, the feature select input of the multiplexer was set to 2. The spatial plus depth feature (See Figure 6a) of the test signature was matched with the feature template and verified using the decision threshold. Quantitatively, Tables 2 and 3 show that number of false rejections and FRR of this verification module are the same as those for the *DSV* module, whereas the number of false acceptances, FAR, and EER are reduced. In summary, Our *3D-SV* module shows the best performance, since it includes complete 3D information altogether, which is inherently present in the in-air signature trajectory.

6.2. Comparison with Other Verification Methods

Since there are no publicly available datasets and codes available for in-air signatures, Table 4 lists the performances of other methods evaluated on their self-built datasets. Alongside, we show the performance of our two best implementations on our self-built dataset. Our *DSV* module shows the competitive performance, whereas the *3D-SV* module shows the best results. It shows that the hidden depth feature in the in-air signature is important for improved performance.

Table 4. The table shows the performances of the existing in-air signature methods and our method. Due to unavailability of a public dataset for in-air signatures, we report results on our dataset. While our *3D-SV* module shows the best results, our *DSV* module, which is based on only depth analysis, shows the competitive performance.

Method	Dataset/Acquisition Method	Result
Nguyen et al. [28]	self-built/Accelerometer	EER: 0.8%
Hasan et al. [1]	self-built/Google glass	Accuracy = 97.5%
Nidal et al. [45]	self-built/ data glove	EER: 2.37%
Jeon et al. [17]	self-built/ depth camera	EER: 0.68%
Moon et al. [46]	self-built/Wifi signal	EER: 4.31%
Yuxun et al. [16]	self-built/RGB camera	FAR: 1.90% and FRR: 2.86%
DSV[**Ours**]	self-built/depth camera	EER: 0.51%
3D-SV[**Ours**]	self-built/depth camera	EER: 0.46%

7. Conclusions and Future Work

In this paper, we presented a real-time automatic in-air signature acquisition and verification framework using a low cost multi-modal depth camera. This paper addresses two major limitations in the existing methods for in-air signature verification. First, given the fact that the existing approaches use heuristic methods for fingertip tracking, which are unstable and impractical, we proposes a new CNN-based hand pose estimation method, which reliably tracks fingertips in real-time. The signature trajectory is recorded using an estimated 3D position of the index fingertip in each depth frame. Second, to explore the potential of the hidden depth feature in the in-air signature trajectory, we created our own dataset, which consists of 600 signatures recorded from 15 different subjects. We investigated the performance of the verification module by performing an ablation study on the spatial and depth features and performed extensive evaluations on our database. Experiments showed that the depth feature itself is sufficient for in-air signature verification. In the future, we plan to extend our database and develop a CNN-based algorithm for in-air signatures classification and matching.

Author Contributions: Conceptualization, J.M. and F.S.; Data curation, J.M.; Software, J.M.; Supervision, A.E., S.A. and D.S.; Writing – original draft, J.M.; Writing – review and editing, A.E., S.A., F.S. and M.I.M.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sajid, H.; Sen-ching, S.C. VSig: Hand-gestured signature recognition and authentication with wearable camera. In Proceedings of the 2015 IEEE International Workshop on Information Forensics and Security (WIFS), Rome, Italy, 16–19 November 2015; pp. 1–6.

2. Bush, L.; Carr, S.; Hall, J.; Saulson, J.; Scott-Simmons, W. Creating a "Classroom of the Future" for P-12 Pre-Service Educators. In Proceedings of the Society for Information Technology & Teacher Education International Conference, Savannah, GA, USA, 7 November 2016; pp. 920–924.

3. Robert, S.N.; Thilagavathi, B. Offline signature verification using support vector machine. In Proceedings of the 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, India, 19–20 March 2015; pp. 1–6.

4. Chandra, S.; Maheskar, S. Offline signature verification based on geometric feature extraction using artificial neural network. In Proceedings of the 2016 3rd International Conference on Recent Advances in Information Technology (RAIT), Dhanbad, India, 3–5 March 2016; pp. 410–414.

5. Bhattacharya, I.; Ghosh, P.; Biswas, S. Offline signature verification using pixel matching technique. *Procedia Technol.* **2013**, *10*, 970–977.

6. Lei, H.; Govindaraju, V. A comparative study on the consistency of features in on-line signature verification. *Pattern Recognit. Lett.* **2005**, *26*, 2483–2489.

7.   Parodi, M.; Gómez, J.C. Legendre polynomials based feature extraction for online signature verification. Consistency analysis of feature combinations. *Pattern Recognit.* **2014**, *47*, 128–140.

8.   Fayyaz, M.; Saffar, M.H.; Sabokrou, M.; Hoseini, M.; Fathy, M. Online signature verification based on feature representation. In Proceedings of the 2015 The International Symposium on Artificial Intelligence and Signal Processing (AISP), Mashhad, Iran, 3–5 March 2015; pp. 211–216.

9.   Van, B.L.; Garcia-Saliceti, S.; Dorizzi, B. On using the Viterbi path along with HMM likelihood information for online signature verification. *IEEE Trans. Syst. Man Cybern. Part B* **2007**, *37*, 1237–1247.

10. Alonso-Fernandez, F.; Fierrez-Aguilar, J.; Ortega-Garcia, J. Sensor interoperability and fusion in signature verification: A case study using tablet pc. In *Advances in Biometric Person Authentication*; Springer: New York, NY, USA, 2005; pp. 180–187.

11. Lee, L.L.; Berger, T.; Aviczer, E. Reliable on-line human signature verification systems. *IEEE Trans. Pattern Anal. Mach. Intell.* **1996**, pp. 643–647.

12. Iranmanesh, V.; Ahmad, S.M.S.; Adnan, W.A.W.; Malallah, F.L.; Yussof, S. Online signature verification using neural network and pearson correlation features. In Proceedings of the 2013 IEEE Conference on Open Systems (ICOS), Kuching, Malaysia, 2–4 December 2013; pp. 18–21.

13. Gruber, C.; Gruber, T.; Krinninger, S.; Sick, B. Online signature verification with support vector machines based on LCSS kernel functions. *IEEE Trans. Syst. Man Cybern. Part B* **2010**, *40*, 1088–1100.

14. Martens, R.; Claesen, L. On-line signature verification by dynamic time-warping. In Proceedings of the 13th International Conference on Pattern Recognition, Vienna, Austria, 25–29 August 1996; Volume 3, pp. 38–42.

15. Feng, H.; Wah, C.C. Online signature verification using a new extreme points warping technique. *Pattern Recognit. Lett.* **2003**, *24*, 2943–2951.

16. Fang, Y.; Kang, W.; Wu, Q.; Tang, L. A novel video-based system for in-air signature verification. *Comput. Electr. Eng.* **2017**, *57*, 1–14.

17. Jeon, J.H.; Oh, B.S.; Toh, K.A. A system for hand gesture based signature recognition. In Proceedings of the 2012 12th International Conference on Control Automation Robotics & Vision (ICARCV), Guangzhou, China, 5–7 December 2012; pp. 171–175.

18. Bailador, G.; Sanchez-Avila, C.; Guerra-Casanova, J.; de Santos Sierra, A. Analysis of pattern recognition techniques for in-air signature biometrics. *Pattern Recognit.* **2011**, *44*, 2468–2478.

19. Khoh, W.H.; Pang, Y.H.; Teoh, A.B.J. In-air hand gesture signature recognition system based on 3-dimensional imagery. *Multimed. Tools Appl.* **2018**, 1–25; doi:10.1007/s11042-018-6458-7.

20. Zhou, X.; Wan, Q.; Zhang, W.; Xue, X.; Wei, Y. Model-based deep hand pose estimation. *arxiv* **2016**, arXiv:1606.06854.

21. Malik, J.; Elhayek, A.; Stricker, D. Simultaneous Hand Pose and Skeleton Bone-Lengths Estimation from a Single Depth Image. In Proceedings of the 2017 International Conference on 3D Vision (3DV), Qingdao, China, 10–12 October 2017.

22. Yuan, S.; Garcia-Hernando, G.; Stenger, B.; Moon, G.; Chang, J.Y.; Lee, K.M.; Molchanov, P.; Kautz, J.; Honari, S.; Ge, L.; et al. Depth-Based 3D Hand Pose Estimation: From Current Achievements to Future Goals. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, Utah, 19–21 June 2018.

23. Kumar, A.; Bhatia, K. A survey on offline handwritten signature verification system using writer dependent and independent approaches. In Proceedings of the 2016 2nd International Conference on Advances in Computing, Communication, & Automation (ICACCA) (Fall), Bareilly, India, 30 September–1 October 2016; pp. 1–6.

24. Yadav, M.; Kumar, A.; Patnaik, T.; Kumar, B. A survey on offline signature verification. *Int. J. Eng. Innov. Technol. Vol.* **2013**, *2*, 337–340.

25. Dalal, S.; Jindal, U. Performance of integrated signature verification approach. In Proceedings of the 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 16–18 March 2016; pp. 3369–3373.

26. Katagiri, M.; Sugimura, T. Personal authentication by free space signing with video capture. In Proceedings of the 5th Asian Conference on Computer Vision, Melbourne, Australia, 3–25 January 2002; Volume 6.

27. Takeuchi, A.; Manabe, Y.; Sugawara, K. Multimodal soft biometrie verification by hand shape and handwriting motion in the air. In Proceedings of the 2013 International Joint Conference on Awareness

Science and Technology & Ubi-Media Computing (iCAST 2013 & UMEDIA 2013), Aizu-Wakamatsu, Japan, 2–4 November 2013; pp. 103–109.

28. Diep, N.N.; Pham, C.; Phuong, T.M. SigVer3D: Accelerometer Based Verification of 3-D Signatures on Mobile Devices. In *Knowledge and Systems Engineering*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 353–365.

29. Matsuo, K.; Okumura, F.; Hashimoto, M.; Sakazawa, S.; Hatori, Y. Arm swing identification method with template update for long term stability. In *International Conference on Biometrics*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 211–221.

30. Creative. Senz3D Interactive Gesture Camera. 2018. Available online: https://asia.creative.com/p/web-cameras/creative-senz3d (accessed on 7 November 2018).

31. Guo, H.; Wang, G.; Chen, X.; Zhang, C.; Qiao, F.; Yang, H. Region Ensemble Network: Improving Convolutional Network for Hand Pose Estimation. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017.

32. Tompson, J.; Stein, M.; Lecun, Y.; Perlin, K. Real-time continuous pose recovery of human hands using convolutional networks. *ACM Trans. Gr.* **2014**, *33*, 169.

33. Tang, D.; Jin Chang, H.; Tejani, A.; Kim, T.K. Latent regression forest: Structured estimation of 3d articulated hand posture. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 24–27 June 2014; pp. 3786–3793.

34. Sun, X.; Wei, Y.; Liang, S.; Tang, X.; Sun, J. Cascaded hand pose regression. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 824–832.

35. Wang, G.; Chen, X.; Guo, H.; Zhang, C. Region Ensemble Network: Towards Good Practices for Deep 3D Hand Pose Estimation. *J. Vis. Commun. Image Represent.* **2018**, *55*, 404–414.

36. Oberweger, M.; Wohlhart, P.; Lepetit, V. Training a feedback loop for hand pose estimation. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 3316–3324.

37. Malik, M.I.; Liwicki, M. From terminology to evaluation: Performance assessment of automatic signature verification systems. In Proceedings of the 2012 International Conference on Frontiers in Handwriting Recognition, Bari, Italy, 18–20 September 2012; pp. 613–618.

38. Nguyen, V.; Blumenstein, M.; Muthukkumarasamy, V.; Leedham, G. Off-line signature verification using enhanced modified direction features in conjunction with neural classifiers and support vector machines. In Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR 2007), Parana, Brazil, 23–26 September 2007; Volume 2, pp. 734–738.

39. Malik, M.; Liwicki, M.; Alewijnse, L.; Ohyama, W.; Blumenstein, M.; Found, B. Signature verification and writer identification competitions for on-and offline skilled forgeries (sigwicomp2013). In Proceedings of the 12th International Conference on Document Analysis and Recognition, Washigton, DC, USA, 25–28 August 2013.

40. Guerbai, Y.; Chibani, Y.; Abbas, N. One-class versus bi-class SVM classifier for off-line signature verification. In Proceedings of the 2012 International Conference on Multimedia Computing and Systems, Tangier, Morocco, 10–12 May 2012; pp. 206–210.

41. Bergamini, C.; Oliveira, L.S.; Koerich, A.L.; Sabourin, R. Combining different biometric traits with one-class classification. *Signal Process.* **2009**, *89*, 2117–2127.

42. Amer, M.; Goldstein, M.; Abdennadher, S. Enhancing one-class support vector machines for unsupervised anomaly detection. In Proceedings of the ACM SIGKDD Workshop on Outlier Detection and Description, Chicago, IL, USA, 11 August 2013; pp. 8–15.

43. Manevitz, L.; Yousef, M. One-class document classification via neural networks. *Neurocomputing* **2007**, *70*, 1466–1481.

44. Sanguansat, P. Multiple multidimensional sequence alignment using generalized dynamic time warping. *WSEAS Trans. Math.* **2012**, *11*, 668–678.

45. Kamel, N.S.; Sayeed, S.; Ellis, G.A. Glove-based approach to online signature verification. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 1109–1113.

46. Moon, H.C.; Jang, S.I.; Oh, K.; Toh, K.A. An In-Air Signature Verification System Using Wi-Fi Signals. In Proceedings of the 2017 4th International Conference on Biomedical and Bioinformatics Engineering, Seoul, Korea, 12–14 November 2017; pp. 133–138.