A Novel Signature Watermarking Scheme for Identity Protection

Sunpreet Sharma
School of Engineering
Western Sydney University
Sydney, Australia
sunpreet.sharma@westernsydney.edu.au

Ju Jia Zou
School of Engineering
Western Sydney University
Sydney, Australia
j.zou@westernsydney.edu.au

Gu Fang
School of Engineering
Western Sydney University
Sydney, Australia
g.fang@westernsydney.edu.au

***Abstract*—A novel non-blind watermarking technique for identity protection is presented. The proposed watermarking scheme uses the owner's signature as the watermark through which the ownership and validity of the document can be proven and kept intact. The proposed scheme is robust, imperceptible and faster in comparison to the other state of the art methods. Experimental simulations and evaluations of the proposed method show excellent results from both objective and subjective view points.**

***Index Terms*—Watermark; Cybersecurity; Wavelets; DWT; Identity**

I. Introduction

Since the dawn of the internet, cybersecurity has played a vital role. However, it has never been as important as it is in today's world. As more and more people are trading/working from home, the internet usage is at its peak. Consequently, this is the prime time for hackers and they are finding novel ways of identity stealing that can later be used for their own advantage at the expense of others. One of the recent examples of identity theft is the Facebook security breach that affected 50 millions users all across the world [1]. Secondly, approximately 150 Australians had their COVID-19 Early Superannuation Release funds stolen due to identity crisis [2]. Lastly, a recent survey by the Australian Department of Foreign Affairs and Trade (DFAT) has shown that almost 3-4% of total passports checked in airports are recognised as fake, all across the world. These are just a few of the examples of a never ending list of such mishappenings with huge implications to individuals and society as a whole. Subsequently, the aim of this paper is to present a watermarking technique with its potential applications in safeguarding the identity document(s) as watermarking is a tool that can certainly limit if not eliminate their unauthorised production.

Watermarking is a process of adding information to a medium (Host signal) in a manner that later on the added information can be extracted in order to verify the medium [3]. A successful extraction proves the validation of the medium and vice-versa. A watermarking scheme depending on the extraction procedure can either be blind or non-blind. For instance, if the host signal is required at the time of extraction it is considered as non-blind else it is blind. However, in the case of a passport it is non-blind because the embedded watermark generally consists of the owner's information, in the form of a signature, fingerprint, iris information etc [4]. These features are matched against those embedded within the document to determine its ownership. A positive correlation among the former and latter validates the ownership, whereas, a mismatch alarms the border control of a country [5].

Presently, identity documents are divided into two categories i.e. e-documents and printed ones. However, efforts are being put into transforming the majority of printed documents into their electronic versions. That being said, it does not mean that the physical verification methods are not required. For instance, in 2019 the Sydney Airport had to stop its operations due to a technical failure in the smart gates and e-passport verification machines [6]. This resulted in a number of passengers being stuck in the airport for multiple hours. One way of resolving such issues is to embed a combination of watermarks (both electronic and physical) in an identity document, so that any of these can be used for verification when one fails. This paper presents a signature watermarking technique that uses the owner's signatures as the watermark.

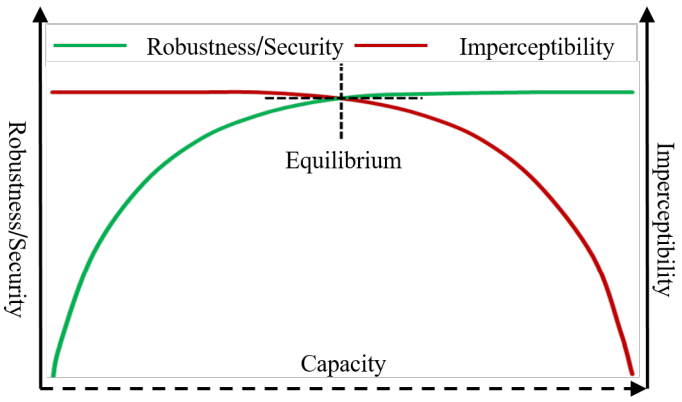

Fig. 1. Existing correlations and trade-offs in watermarking.

A successful watermarking technique needs to fulfil three main correlated requirements known as imperceptibility, security/robustness and capacity [3]. Fig. 1 illustrates that the higher the amount of embedded information i.e. capacity, the better is the watermarking scheme. However, the result of adding a large amount of information can be detrimental

to imperceptibility, whereas, a small amount can improve the imperceptibility but degrades the robustness and security. Consequently, an efficient method to find the equilibrium between these trade-offs is a major challenge in the field of watermarking [3]. The proposed watermarking scheme successfully addresses the aforementioned challenge and the main contributions of the proposed work are mentioned as below.

- A non-blind dual watermarking scheme based on Discrete Wavelet Transform (DWT) is presented. The use of DWT in the proposed method is favourable due to its excellent spatial-frequency localization [7]. The novel DWT coefficient block selection process not only makes it faster in processing but also empowers the proposed watermarking scheme with a high degree of security and robustness.
- The proposed watermark embedding is imperceptible and the watermarked image gives excellent results in terms of Peak Signal to Noise Ratio (PSNR) values.
- Working illustrations of the proposed watermarking scheme are demonstrated and tested against both geometrical and non-geometrical attacks (see [3] for an insight on various attacks).

The rest of this paper is organised in the following manner. Section II covers the related work in the field, Section III discusses the proposed methodology, Section IV and Section V contain experimental results and the conclusion.

II. Related Work

Digital watermarking is highly commendable in achieving goals such as authentication, copyright protection and security of the multimedia data. Moreover, it is an active area of research and this section highlights the main state of the art works which are inline with the proposed watermarking scheme. Lin et al. developed a DWT coefficient difference based watermarking technique [8]. This approach is widely adopted and influenced the later works such as [9], [10]. Despite the method's success in achieving very high imperceptibility and decent capacity, it struggled from the security aspect [11]. Subsequently, to address this shortfall, an idea to embed multiple watermarks is one of the adopted remedies. This ideology is successfully being practiced within the most recent literature on watermarking for identity (ID) protection or similar applications [4], [5], [12]. Furthermore, the disadvantages associated with methods that are solely based on DWT are restricted if not nullified by pairing it with other techniques such as Discrete Cosine Transform (DCT) [13]–[17]. However, embedding multiple watermarks is a tedious task, has its own challenges and it is out of the scope of this paper. The proposed work is positively influenced by the techniques used in [8]–[10] and subsequently bridges its aforementioned gaps.

III. Proposed Methodology

Fig. 2 gives an overview of the proposed watermarking scheme. Firstly, the host image/original signal of size nxn (512x512 in this paper) is decomposed into wavelet coefficients by using DWT [8]. It is crucial to justify that alike the methods in [8]–[10], a three level decomposition using DWT is selected in the proposed method, so that a fair comparison between the two can be established. In Fig. 3, a three level decomposition of the host image (represented by burgundy colour) extracts ten subbands, consisting of one approximate subband $LL3$ (represented by solid green) and nine detail subbands (LH, HH, HL) as a result of lowpass and highpass filtering, respectively. It is well known that the approximate subband is comprised of low-frequency components and it contributes to the majority of the image information and can simply lead to distortions when altered, thus, it is not preferable for watermark embedding. Subsequently, embedding a watermark in $HL3$ and $HH3$ subbands (represented by solid orange and blue, respectively) is also not suitable as the watermark can easily be breached by image processing attacks such as rotation, lossy compression etc [3]. Based on the aforementioned reasons, $LH3$ subband (represented by solid yellow coloured block in Fig. 3) is selected for signature watermark embedding. A more detailed analysis of watermark embedding in each of these subbands can be found in [10]. The rest of the proposed watermarking strategy is covered within the upcoming subsections on watermark embedding in this paper.

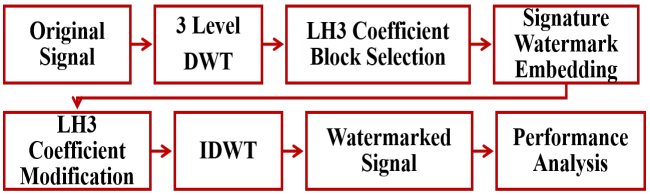

Fig. 2. Blueprint of the proposed method.

A. Signature Watermark Embedding

The proposed work varies in the embedding block selection procedure(s) in a number of ways to the one adopted by authors in [8]–[10].

First and foremost, methods in [8]–[10] use a number of secret keys (three at a given instance). Notwithstanding the success of using multiple keys in achieving robustness, they tend to require a high computational power, whereas the proposed method uses only one secret key throughout. Secondly, the former is not optimal for transmission as all of these keys are required at the time of watermark extraction and this can cause an issue where transmission bandwidth is limited, thus, adversely impacting the capacity. However, in the proposed method, as only one key being used, it makes the transmission process faster and the watermark with a higher payload/size can be inserted, thus improving its capacity. Finally, the coefficient block selection procedure is performed through shuffling at multiple occasions using multiple keys in [8]–[10], in order to achieve security/robustness. In the proposed method this aspect is bridged by using only one

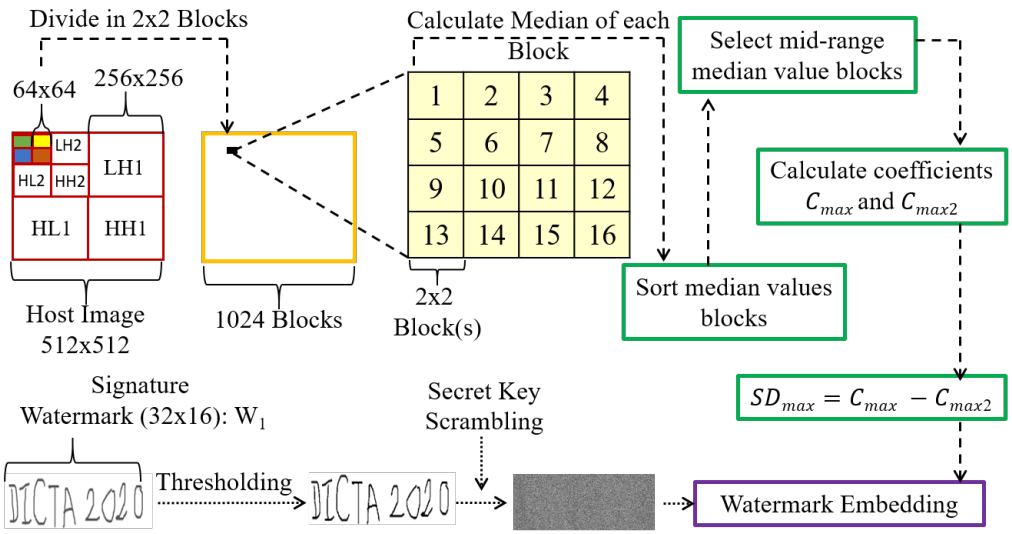

Fig. 3. Signature watermark embedding process. Digits within the light yellow coloured block(s) correspond to a 2x2 block.

security key in conjunction with the novel DWT coefficient block selection for watermark embedding.

Inspired from [8]–[10] and for aforementioned reasons $LH3$ subband of size 64x64 (4096 bits) is selected in this process. In Fig. 3, $LH3$ coefficients are divided into a total of 1024 non-overlapping blocks, each is 2x2 in size. The median of each of these blocks is calculated and are sorted either in an ascending or descending order (see [15] for median calculations). Furthermore, 512 of these median values falling within the mid-range of total 1024 values are selected and so are the 512 corresponding 2x2 blocks. The selected 512 blocks are termed as M_b blocks in the proposed method. Subsequently, the difference among the maximum valued (C_{max}) and the second maximum valued (C_{max2}) coefficient is calculated. This difference is termed as *Maximum Significant Difference* (SD_{max}) and this process is repeated for each of the 512 selected M_b blocks.

Moreover, the signature watermark (W_1) of size 32x16 (512 bits) is prepared by following the same series of steps as performed above in Fig. (3). The very first step involves the binarization of the selected watermark by thresholding it to a value of 128. This limits the watermark to only two pixel vales of 0 (Black) and 255 (White) which corresponds to 0 and 1 in binary, respectively. Subsequently, the binarized watermark is scrambled by using a secret key. The purpose of using such key is to maintain the integrity of the medium during transmission because the very same key is required at the time of validation which is achieved through the watermark extraction process (discussed later in this paper). The Fisher–Yates shuffle algorithm is used throughout this paper due to its robust performance and state of the art usage (see [9] for more details on this shuffling concept). Thereafter, embedding is initiated by an embedding quantizer (EQ). For each watermark bit the EQ (represented by the purple boundary in Fig.3) quantizes the highest valued coefficient (C_{max}) of a corresponding block to C_{max}^{new} by using the Equations (1) and (2). If the watermark bit is 1,

$$C_{max}^{new} = \begin{cases} C_{max} + T, & if \ \ (SD_{max}) < max(\lambda, T) \\ C_{max}, & otherwise \end{cases} \tag{1}$$

and if the watermark bit is 0, the corresponding significant difference is directly quantized to zero by equalising the maximum and second maximum coefficient.

$$C_{max}^{new} = C_{max2} \tag{2}$$

In Equation (1), λ is the *average significant difference* of all the M_b blocks, calculated as per Equation (3) and T is the threshold value used for quantization. Similar to [9], $T = 11$ is the chosen threshold value for this paper as well, whereas a complete account on the empirical selection of T is given in [8]. The floor function in Equation (3) is represented by $\lfloor . \rfloor$.

$$\lambda = \left\lfloor \frac{\sum_{i=1}^{M_b} SD_{max}^{i}}{M_b} \right\rfloor \tag{3}$$

Subsequent steps in the proposed watermarking scheme includes the merging of modified coefficient blocks with unaltered blocks of *LH3* subband and performing inverse DWT (IDWT), respectively. Consequently, a final watermarked image comprised of the owner's signature as the watermark is achieved by using Equation (4).

$$WI_{Final} = HI_{Original}(1 + \alpha W_{Total}) \tag{4}$$

where WI_{Final}, $HI_{Original}$, W_{Total} and α stand for the final watermarked image, original host image, total watermark embedded and the watermark strength parameter, respectively. The value of α ranges between (0 1] and it defines the visibility of the inserted watermark with "1" being fully visible and vice versa [18]. Alike [9] and [10], α equal to 0.6 is chosen for experimental simulations in this paper.

B. Watermark Extraction

A non-blind watermarking requires both host signal and watermarked signal at the time of extraction. Thus, the extraction process used in the proposed non-blind watermark extraction is defined as in Equation (5).

$$W_{Total} = \frac{WI_{Final} - HI_{Original}}{\alpha HI_{Original}} \tag{5}$$

It is essential to realise that Equation (5) only outputs the watermark(s) in a scrambled state. The final step in watermark extraction is unscrambling the former by an inverse execution of the aforementioned secret key.

IV. Experimental Results

Fig. 4. Imperceptibility comparisons of the proposed scheme. Test/Host images in the solid blue (left-right:) *Lady, Zelda, Baboon* and dashed blue is for the *original signature watermark*. Solid purple boundary is of the watermarked images without any attack whereas the dashed purple boundary represent the extracted watermarks. Best viewed when zoomed in.

Fig. (4) illustrates the comparisons among the original/host images and the watermarked images without any attacks. Subjectively, it can be noticed that the watermarked image(s) appears to be serene and homogeneous in tone. Consequently, the gained image(s) display a smooth transition between the grey levels and as a result, the achieved watermarked image is imperceptible to the Human Visual System (HVS). Furthermore, Fig. 5 shows the watermarked images under various attacks such as rotation attack at $45°$, Gaussian Noise at 0.001 and JPEG compression Quality Factor (QF=40) and their respective extracted watermarks. The similarity between the embedded and extracted watermarks is calculated by Normalized Correlation (NC) from Equation (7).

Table I contains a quantitative evaluation of the proposed method in terms of imperceptibility, robustness/security and capacity in which a watermarked image attained by the proposed method is compared to other state of the art methods such as [14], [15], [17]. The imperceptibility is measured in decibels (dB) through Peak Signal to Noise Ratio (PSNR)

Fig. 5. Robustness/Security comparisons of the proposed scheme. Solid red, yellow and green boundaries contain the watermarked images under rotation attack at $45°$, Gaussian Noise at 0.001 and JPEG compression (QF=40), respectively. All dashed boundaries represent the extracted watermarks from the attacked watermarked images. Best viewed when zoomed in.

given by Equation (6). *A high PSNR value indicates high imperceptibility*.

$$PSNR = 10\log_{10}\frac{(2^d-1)^2 BH}{\sum_{i=1}^{B}\sum_{j=1}^{H}(z[i,j]-p[i,j])^2} \tag{6}$$

where d is the bit depth of a pixel, B and H are the image breadth and height, respectively. Furthermore, $z(i,j)$ and $p(i,j)$ indicate pixel values of the host/original image (*without* any watermark) and the watermarked image produced as a result of the dual watermarking operation, respectively. Subsequently, the robustness of the proposed method is tested through Normalized Correlation (NC) by Equation (7).

$$NC = \frac{\sum_{i=1}^{P_w}\sum_{j=1}^{Q_w}(W[i,j]\times W^{'}[i,j])}{\sqrt{\sum_{i=1}^{P_w}\sum_{j=1}^{Q_w}(W^2[i,j])}\times\sqrt{\sum_{i=1}^{P_w}\sum_{j=1}^{Q_w}(W'^2[i,j])}} \tag{7}$$

where W and $W^{'}$ stands for the original and extracted watermarks of dimensions P_w and Q_w, respectively. NC values ought to have a range between [0 1], *with* 0 *being least in similarity and* 1 *as the highest*. The versatility of the proposed watermarking scheme is tested on 110 greyscale images (available at [19]) and the average of test results are summarised in Table I. These experiments are conducted on a machine with i7-8650U CPU running at 1.9 GHz, 16 GB RAM and 64-bit operating system using MATLAB (R2018a).

TABLE I
PERFORMANCE ANALYSIS.

Methods / Attacks	**PSNR (dB)**	**Time (Seconds)**	**NC (W_1)**		
Method [9] / No Attack	41.89	5.89	0.94	0.95	0.97
Method [10] / No Attack	42.95	6.01	0.97	0.97	0.96
***Ours** / No Attack*	**43.68**	**5.13**	**0.98**	**0.98**	**0.99**
Method [9] / Rotation	40.76	6.32	0.92	0.94	0.95
Method [10] / Rotation	42.06	6.59	0.95	0.95	0.94
***Ours** / Rotation*	**42.86**	**5.51**	**0.96**	**0.97**	**0.98**
Method [9] / Gaussian Noise	40.8	6.26	0.94	0.93	0.91
Method [10] / Gaussian Noise	42.34	6.48	0.96	0.96	0.92
Ours / Gaussian Noise	**42.98**	**5.42**	**0.97**	**0.98**	**0.98**
Method [9] / Compression	40.74	6.59	0.91	0.9	0.92
Method [10] / Compression	41.67	6.84	0.93	0.92	0.94
***Ours** / Compression*	**42.46**	**5.64**	**0.95**	**0.94**	**0.95**

Table I exhibits that the method in [10] is better in imperceptibility and robustness in comparison to the method in [9], however, the latter is faster in the processing time (measured in seconds). The proposed watermarking scheme outperforms both of these methods in terms of imperceptibility and robustness. Furthermore, it performs at par with the other two existing methods in terms of capacity as each of these methods are able to embed 512 bits of watermark while maintaining an ample trade-off between imperceptibility and robustness. Lastly, the proposed methods is faster in processing than both of its counterparts, thus, making it superior and feasible from its application view point. W_1 in Table I depicts the signature watermark. It is to be noted that the PSNR and the time (in seconds) entries in Table I represent the averaged values of 110 watermarked test images under a certain attack or no attack, respectively.

V. CONCLUSION

This paper presents a novel signature watermarking approach for identity protection. Firstly, the proposed watermarking scheme uses the owner's signature as the watermark through which ownership and validity of the document can be proven and kept intact. Consequently, in terms of capacity, the proposed scheme is at par with the existing state of the art methods. Secondly, the proposed embedding approach uses both Discrete Wavelet Transform (DWT) and a novel median based embedding block selection method. Working in conjunction together, these two techniques enable the proposed watermarking scheme to outperform the existing methods in terms of imperceptibility and robustness/security.

ACKNOWLEDGMENT

This work is supported by the Western Sydney University Postgraduate Research Award. Thanks to Jessica Johnston for proof-reading this work.

REFERENCES

[1] K. van der Schyff, S. Flowerday, and S. Furnell, "Duplicitous social media and data surveillance: An evaluation of privacy risk," *Computers & Security*, p. 101822, 2020.

[2] An australian couple almost had $20,000 of superannuation stolen. the scam has exposed serious flaws in the federal government's early access scheme. Business Insider Australia. Sydney, Australia. [Online]. Available: https://www.businessinsider.com.au/early-access-super-scheme-scam-withdrawal-fraud-ato-mygov-2020-5

[3] S. Sharma, J. J. Zou, and G. Fang, "Recent developments in halftone based image watermarking," in *2019 International Conference on Electrical Engineering Research & Practice (ICEERP)*. IEEE, 2019, pp. 1–6.

[4] K. Bobkowska, K. Nagaty, and M. Przyborski, "Incorporating iris, fingerprint and face biometric for fraud prevention in e-passports using fuzzy vault," *IET Image Processing*, vol. 13, no. 13, pp. 2516–2528, 2019.

[5] L. R. Haddada, B. Dorizzi, and N. E. B. Amara, "A combined watermarking approach for securing biometric data," *Signal Processing: Image Communication*, vol. 55, pp. 23–31, 2017.

[6] Airport delays subside following passport machine failures. ABC NEWS AUSTRALIA. Sydney, Australia. [Online]. Available: https://www.abc.net.au/news/2019-07-15/sydney-international-airport-delays-passport-control/11309132

[7] S. A. Parah, J. A. Sheikh, N. A. Loan, and G. M. Bhat, "Robust and blind watermarking technique in dct domain using inter-block coefficient differencing," *Digital Signal Processing*, vol. 53, pp. 11–24, 2016.

[8] W.-H. Lin, S.-J. Horng, T.-W. Kao, P. Fan, C.-L. Lee, and Y. Pan, "An efficient watermarking method based on significant difference of wavelet coefficient quantization," *IEEE Transactions on Multimedia*, vol. 10, no. 5, pp. 746–757, 2008.

[9] V. S. Verma, R. K. Jha, and A. Ojha, "Significant region based robust watermarking scheme in lifting wavelet transform domain," *Expert Systems with Applications*, vol. 42, no. 21, pp. 8184–8197, 2015.

[10] M. Islam, A. Roy, and R. H. Laskar, "Svm-based robust image watermarking technique in lwt domain using different sub-bands," *Neural Computing and Applications*, vol. 32, no. 5, pp. 1379–1403, 2020.

[11] P. Meerwald, C. Koidl, and A. Uhl, "Attack on "watermarking method based on significant difference of wavelet coefficient quantization"," *IEEE transactions on multimedia*, vol. 11, no. 5, pp. 1037–1041, 2009.

[12] M. Barr and C. Serdean, "Wavelet transform modulus maxima-based robust logo watermarking," *IET Image Processing*, vol. 14, no. 4, pp. 697–708, 2019.

[13] X.-b. Kang, F. Zhao, G.-f. Lin, and Y.-j. Chen, "A novel hybrid of dct and svd in dwt domain for robust and invisible blind image watermarking with optimal embedding strength," *Multimedia Tools and Applications*, vol. 77, no. 11, pp. 13 197–13 224, 2018.

[14] N. N. Hurrah, S. A. Parah, N. A. Loan, J. A. Sheikh, M. Elhoseny, and K. Muhammad, "Dual watermarking framework for privacy protection and content authentication of multimedia," *Future Generation Computer Systems*, vol. 94, pp. 654–673, 2019.

[15] N. A. Loan, N. N. Hurrah, S. A. Parah, J. W. Lee, J. A. Sheikh, and G. M. Bhat, "Secure and robust digital image watermarking using coefficient differencing and chaotic encryption," *IEEE Access*, vol. 6, pp. 19 876–19 897, 2018.

[16] A. K. Abdulrahman and S. Ozturk, "A novel hybrid dct and dwt based robust watermarking algorithm for color images," *Multimedia Tools and Applications*, vol. 78, no. 12, pp. 17 027–17 049, 2019.

[17] A. K. Singh, B. Kumar, S. K. Singh, S. Ghrera, and A. Mohan, "Multiple watermarking technique for securing online social network contents using back propagation neural network," *Future Generation Computer Systems*, vol. 86, pp. 926–939, 2018.

[18] D. Bhowmik and C. Abhayaratne, "Embedding distortion analysis in wavelet-domain watermarking," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 15, no. 4, pp. 1–24, 2019.

[19] Usc-sipi image databases. University of Southern California. California, USA. [Online]. Available: http://sipi.usc.edu/database/