plates, especially due to requirements such as fighting numerous occurrences of crimes in countries borders. Our solution compensates the removal of the country/region information from the license plates, in such a way that we do not need an online system to search in a vehicles database.

3.2. Improving OCR performance

Based on data provided by the Brazilian Federal Highway Police (PRF, which stands for *Polícia Rodoviária Federal*), we can define an approach to improve OCR performance, in addition to other well-known configurations and procedures, as described in **Error! Reference source not found.** or OCR vendor manuals.

The PRF traffic oversight activities counts on hundreds of cameras spots in different regions of Brazil. The basic setup consists of cameras to capture license plates imagens from the front and the rear of a vehicle. Due to poor light conditions, rainy weather in some regions, dirty license plates and illegible characters, the OCR performance can be as bad as 60% of correctness. The correctness of the OCR cascade was admeasured based on 10000-labeled images, validated by humans, from 5000 vehicles with two license plates each. All images were chosen at random from different camera sets in different locations and time of day.

After the human validation of license plates images, it was possible to verify which are the most confusing characters and its positions in the sequence of characters.

The new Mercosul license plates characters' font is different from the current plates in use in Brazil. The chosen font for the new license plates model is the *FE Engschrift* font, which is designed to make characters' forgery more difficult and to improve OCR systems readings. Even so, the characters that are more frequently misread can be avoided when generating a license plate identification string. Then, for example, if the first character of a newly generated plate is a D, the generator system can check in the existing plates database whether there exists a plate with a character in the first position that is frequently identified incorrectly as a D, such as a 0 or a O.

We can avoid the generation of plates whose characters may be misread due to the existence of another plate with confusing characters in the same position. However, it may not be possible to exclude a big set of license plates, but the plates generator can postpone the generation of plates with confusing characters as long as possible. In addition, it may be helpful to issue license plates with confusing characters to vehicles in different and distant regions. In doing that, it may be possible to infer automatically the characters of a license plate in case of OCR misreading, based on previous readings of that plate in a given region and time window or by checking it in a central database. This is a complementary strategy to improve LPR stations performance and it can be used along with those proposed in **Error! Reference source not found.**, **Error! Reference source not found.** and **Error! Reference source not found.**, for example.

If an online database with all vehicles license plates is available, we can propose an improvement to automatically identify uniquely a vehicle, even when the LPR station system make an incorrect OCR reading. A possible approach is to filter generated plates identification strings and to use the confusion matrix idea from **Error! Reference source not found.**. Each element in the matrix represents the conditional probability $p(r|t)$ of a true character t being recognized as r. On generating a license plate p^{new}, we compare each character of p^{new} with characters in corresponding positions of all existing plates p^{DB} in a central database, which contains all already issued license plates. Consider a character r_i in position $i: 0 < i \leq |p^{\text{new}}|$ of the newly generated license plate. Also, consider a character t_i of any plate p^{DB} from the database. Then, we interpret $p(r_i|t_i)$ as the probability of having the character t_i from p^{DB} being recognized as r_i, which may increase the total probability of a false positive when searching p^{DB} in the database after p^{new} is accepted as a new license plate in the database. Summing up probabilities $p(r_i|t_i)$ for all positions i, we accept a new generated license plate p^{new} in the database whenever $\sum_i p(r_i|t_i) < \mu$, where μ is an arbitrary system parameter.

Furthermore, we can consider that many LPR stations operate with a magnetic loop detector capable of identifying some types of vehicles, such as cars, buses, trucks and motorcycles. Then, with this additional information, we can issue a plate with $\sum_i p(r_i|t_i) > \mu$ to vehicles of different categories.

4. License plates generation method

We propose a method for vehicle plates generation, inspired on the ideas proposed for Personal Identification Numbers in **Error! Reference source not found.**. We are using the same method to get a decentralized unique pseudorandom numbers generation, based on some public distributed information of a setup phase.

4.1. Defining a domain in $\mathbb{N}$ for license plates

Let us establish a representation of license plates as natural numbers. Define an injective and invertible function $f: P \to N$ for any given plate $p \in P$, such that $f(p) = n$, $n \in N \subset \mathbb{N}^*$ and $|P| = |N|$. $|P|$ is the domain of all possible license plates issued by an authority A. Now, we use a positional base-36 numeral system with symbols (digits) $0 \dots 9$, $A \dots Z$. The value of $|P|$ depends on the policies defined for the license plates generation system, which is usually a decision of a central authority or a designated workgroup.

We will consider that all symbols can be combined in any position to form a license plate, resulting in $|P| = 36^m$, where m is the number of symbols in a license plate. Besides, each plate is generated by an authority A_i as a unique pseudorandom natural number in a decentralized environment, where $0 < i \leq a$ and a is the number of authorities engaged on a common decentralized vehicle registration system.

4.2. Generating license plates

The intuition for the plates generation is that an authority A_i, $0 < i \leq a$, gets a set Γ_i of prime numbers used to generate plates $p_j^i \in P_i$, represented by natural numbers in base-36 $p_j^{i_{36}} \in N_i \subset \mathbb{N}$, $j \in \mathbb{N}: j \leq |P_i|$. Under the *Fundamental Theorem of Arithmetic*, we select all the prime numbers $\gamma_r^i \in \Gamma_i, 0 < r \leq |\Gamma_i|$ and natural exponents $\zeta_1^i, \zeta_2^i, \dots, \zeta_{|\Gamma_i|}^i \in \mathbb{N}$ in order to generate the natural number $p_j^{i_{36}} \in N_i$.

$$p_j^{i_{36}} = \prod_{x=1}^{|\Gamma_i|} (\gamma_x^i)^{\zeta_x^i} = (\gamma_1^i)^{\zeta_1^i} . (\gamma_2^i)^{\zeta_2^i} \dots (\gamma_{|\Gamma_i|}^i)^{\zeta_{|\Gamma_i|}^i} \quad (1)$$

For example, for plates of 7 characters, the last plate is, lexografically, $p_j^i = ZZZZZZZ \in P_i$, and $f(ZZZZZZZ) = p_j^{i_{36}} = 78\,364\,164\,095$, for some authority $\mathrm{A_i}$, where the factors of $p_j^{i_{36}}$ are in Γ_i, i.e., $\{5,7,29,197,55987\} \subset \Gamma_i$. The greatest factor of $p_j^{i_{36}}$ defines the authority A_i, that is, no other authority $A_k, 0 < k \leq a, i \neq k$, can issue a plate with the greatest factor equals to 55987. We say that the factor $\gamma_x^i = 55987$ is an element of the *signature set* Γ_i^{sig}, $55987 \in \Gamma_i^{\text{sig}}$, of the authority A_i, such that $\Gamma_i^{\text{sig}} \subset \Gamma_i$.

In order to generate a random plate $p_j^{i_{36}}$ as a random factored number of equation (1), we can use the efficient algorithm presented in **Error! Reference source not found.**. The algorithm generates a number n uniformly distributed over $\{1, \dots, m\}$, where m represents the natural number representation of the last possible plate, considering a lexicographical order.

4.3. Setup Phase

The setup phase consists of distributing the public set of prime numbers Γ_i for each authority A_i. Be authorities A_i and A_k, $i \neq k$, with expected license plates domains P_i and P_k, respectively. If $|P_i| > |P_k|$, then A_i gets Γ_i capable of generating more plates than A_k.

Reference **Error! Reference source not found.** describes a method to construct the prime numbers set Γ_i for an authority A_i, based on the expected size of $|P_i|$. With a the number of authorities in the system, the method iteratively constructs $\Gamma_1, \Gamma_2, \dots, \Gamma_a$, given $|P_1| > |P_2| > \dots > |P_a|$.

For any two distinct authorities A_i and A_{i+1}, the following holds $|P_i| > |P_{i+1}|$ and $\Gamma_\mathrm{i} \subset \Gamma_{i+1}$. Define $\Gamma_{i+1} = (\Gamma_i \cup \Gamma_{i+1}^{\text{sig}})$, where $\Gamma_{i+1}^{\text{sig}}$ is a set where all elements (prime numbers) $\gamma_s^{i+1} \in \Gamma_{i+1}^{\text{sig}}$ are greater than any element $\gamma_r^i \in \Gamma_i$, where $0 < s \leq |\Gamma_{i+1}^{\text{sig}}|$ and $0 < r \leq |\Gamma_i|$, i.e.:

$$\forall \gamma_s^{i+1} \in \Gamma_{i+1}^{\text{sig}}, \forall \gamma_r^i \in \Gamma_i : \gamma_s > \gamma_r \quad (2)$$

With sets Γ_i and Γ_{i+1}, authorities A_i and A_{i+1} can generate, at least, $|P_i|$ and $|P_{i+1}|$ plates, respectively. The first prime numbers set Γ_1 for authority A_1 has the smallest prime numbers, $\Gamma_1 = \{2,3,5, \dots\}$. With the smallest primes, A_1 is expected to be the authority who can generate the greatest amount of license plates. Additionally, $\Gamma_1 = \Gamma_1^{\text{sig}}$, that is, any generated license plate $p_j^{1_{36}}$, $0 < j \leq |P_1|$, has the greatest prime factor in $\Gamma_1 = \Gamma_1^{\text{sig}}$.

In order to generate a plate $p_j^{i_{36}}$ for any authority A_i, it must be included as factors of $p_j^{i_{36}}$ one or more prime numbers $\gamma_r^i \in \Gamma_i^{\text{sig}}$, $0 < r < |\Gamma_i^{\text{sig}}|$. That way, each prime γ_r^i has the role of a signature for the authority A_i.

4.4. Decoding the vehicle plates authority

Given any license plate p^y issued by an authority A_y, we need to find $y \in \mathbb{N}^* : y \leq a$.

First, we find $p^{y_{36}} = f(p^y)$. Then, factorize $p^{y_{36}}$ and identify the greatest factor $\gamma_{\max}$. Finally, find the value of y by locating the public set $\Gamma_\alpha^{\text{sig}}$, such that $\gamma_{\max} \in \Gamma_\alpha^{\text{sig}}$ and constant $\alpha: 0 < \alpha \leq a$. Then, output $y = \alpha$.

Summing up, given Γ_i for authority A_i, be any license plate in the base-36 representation $p_j^{i_{36}} = \prod_{x=1}^{|\Gamma_i|} (\gamma_x^i)^{\zeta_x^i}$, $0 < j \leq |\Gamma_i|$. Therefore, be $\Gamma_i^{p_{36}}$ the set of all prime factors of $p_j^{i_{36}}$. Then, the following relations hold:

$$\Gamma_1 = \Gamma_1^{\text{sig}} \quad (3)$$

$$\Gamma_{i+1} = \Gamma_i \cup \Gamma_{i+1}^{\text{sig}}, i < a \quad (4)$$

$$\Gamma_i \subset \Gamma_k \text{ when } |P_i| > |P_k| \quad (5)$$

$$\Gamma_i^{\text{sig}} \cap \Gamma_k^{\text{sig}} = \emptyset, \forall i, k \leq a, i \neq k \quad (6)$$

$$\forall p_j^{i_{36}} \in N_i : \max\left(\Gamma_{\text{i}}^{p_{36}}\right) \in \Gamma_{\text{i}}^{\text{sig}} \quad (7)$$

$$p_j, p_{j'} \in \bigcup_{k=1}^{a} P_k$$
$$\text{with } j, j' \in \mathbb{N}^*, \text{ then } p_j \approx^v p_{j'} \quad (8)$$

We described relations (3), (4) and (5) in section 4.3. (6) means that each authority has its own exclusive signature set. In (7), all base-36 license plates have their greatest factor in the signature set of its own authority. And, in (8), every two plates of any authority are visually indistinguishable from each other.

4.5. Centralized environment

Although the plates' generation system allows a decentralized operation based on a setup information, it can be used in a centralized environment as well, in order to save upfront costs of infrastructure and workers. In this case any plate can still encode an authority and it also can be visually indistinguishable.

5. Conclusions and future work

With a uniform plate system, citizens may experience more privacy when in traffic in regions where they fell some prejudice, for example. The visual indistinguishability is a benefit, even though the authority decoding is an easy process.

We have proposed more control on the vehicle license plates generation, in order to enhance the effectiveness of ALPR systems.

As a future work, we intend to implement a license plate generator as an open source tool.

6. References

[1] Grupo Mercado Común (GMC) (2014), 'Resolución - 033/2014 – Patente e Sistema de Consultas sobre Veículos do Mercosul';http://www.mercosur.int/innovaportal/v/57 58/2/innova.front/resoluciones_2014 (22 August 2016).

[2] Nogueira, R. B. and Puttini, R. S. (2015) 'On Unique Personal Identifiers', Journal of Internet Technology and Secured Transaction 4.1, pp. 366-72.

[3] Oliveira-Neto, F. M., Han, L. D. and Jeong, M. (2009) 'Tracking large trucks in real time with license plate recognition and text-mining techniques', Transportation Research Record: Journal of the Transportation Research Board 2121, pp. 121-127.

[4] Oliveira-Neto, F. M., Han, L. D. and Jeong, M. K. (2013) 'An online self-learning algorithm for license plate matching', IEEE Transactions on Intelligent Transportation Systems 14.4, pp. 1806-1816.

[5] Hargrove, S. R., Lim, H. and Han, L. D. (2016) 'Expanding License Plate Matching Capabilities with Secondary Self-Learning Algorithm', Transportation Research Record: Journal of the Transportation Research Board 2594, pp. 51-60.

[6] Frayssinet, F. (2015), 'Latin American Migrants Suffer Prejudice in Their Own Region', Inter Press Service; http://www.ipsnews.net/2015/02/latin-american-migrants-suffer-prejudice-in-their-own-region/(10 September 2016).

[7] McLeod, S. A. (2008), 'Prejudice and Discrimination'. SimplyPsychology website;http://www.simplypsychology.org/prejudice.html (10 September 2016).

[8] Pettigrew, T. F. (1959) 'Regional differences in anti-Negro prejudice', Journal of abnormal psychology, 59(1), 28.

[9] Batista, J. R., Leite, E. L., Torres, A. R. and Camino, L. (2014) 'Black and northeasterners: similarities in racial and regional stereotypes', Rev. psicol. polít. vol.14 no. 30 São Paulo, ISSN 1519-549X.

[10] Chang, S., Chen, L., Chung, Y. and Chen, S. (2004) 'Automatic License Plate Recognition', IEEE Transactions on Intelligent Transportation Systems, Vol. 5, No. 1.

[11] Du, S. Ibrahim, M., Shehata, M. and Badawy, W. (2013) 'Automatic License Plate Recognition (ALPR): A State of the Art Review', IEEE Transactions on Circuits and Systems for Video Technology, Vol 23, Issue 2.

[12] Elovic, P. and Scheppmann, L. (2002) 'Practicalities of Implementing OCR: A Case Study of the World's Largest Installation', TOC 2002 Americas, Miami, Florida.

[13] Patel, C., Patel, A. and Patel, D. (2012) 'Optical Character Recognition by Open Source OCR Tool Tesseract: A Case Study', International Journal of Computer Applications (0975 – 8887). Vol 55.

[14] OpenALPR Website. http://www.openalpr.com (Jan 29, 2017).

[15] Presse- und Öffentlichkeitsarbeit (2016) 'Kfz-Kennzeichen und auslaufende Kennzeichen in Deutschland', Stand: 28.12.2016; http://www.kba.de/SharedDocs/Publikationen/DE/Presse/kfz_kennzeichenliste_faltblatt_pdf.pdf?__blob=publicationFile&v=33 (30 January 2017).

[16] Oliveira-Neto, F. M., Han, L. D. and Jeong, M. K. (2012) 'Online license plate matching procedures using license-plate recognition machines and new weighted edit distance', Transportation research part C: emerging technologies 21.1, pp. 306-320.

[17] Shoup, V. (2009) 'A Computational Introduction to Number Theory and Algebra', Cambridge University Press; 2 Edition.

7. Acknowledgements

The authors thank the Technological Solutions team of the Brazilian Federal Highway Police for the support and for allowing us to access all statistical data we needed.

