EVALUATION OF LICENSE-PLATE LOCATION ALGORITHMS

Manuel Menezes de Sequeira *Tomás Brandão* *Miguel Albuquerque*

Manuel.Sequeira@iscte.pt Tomas.Brandao@iscte.pt M.S.Albuquerque@netcabo.pt

ABSTRACT

We propose a new evaluation method for license-plate location algorithms. As a first step towards a comparison of the algorithms described in the literature over the years, we have implemented some relevant algorithms with different approaches and compared them using the proposed evaluation method.

The software used for evaluation and a set of annotated evaluation images are both readily available, making the reproduction of the presented results possible. The availability of both software and an evaluation set allows the results presented in this paper to be used as benchmarks when developing plate location algorithms.

1. INTRODUCTION

License-plate recognition (LPR) systems can typically be divided in two main blocks organized in a cascade: a license-plate location (LPL) block and an optical character recognition (OCR) block. The LPL block uses a LPL algorithm, or a combination of LPL algorithms, to perform its function. The performance of the LPL algorithms is critical for the performance of the overall system [1]: although OCR algorithms can often achieve success rates over 99% [6], the success rates of LPL algorithms are often lower (see Section 5).

A wide range of LPL algorithms have been proposed in the literature (e.g., [2, 8, 9, 11]). Most of the papers present performance measures which are supposed to facilitate a comparison of the proposed LPL algorithm to other existing algorithms. However, this comparison is virtually impossible, since not only performance evaluation methods vary from paper to paper, but also the evaluation sets are totally different and often too small. In this situation, a possible solution for the researcher in the area would be to implement the algorithms in the literature and to test them under fair and equal conditions. However, this task is also bound to fail, since most published algorithms are underspecified. This last problem is almost impossible to solve unless researchers adopt the Clarbout principle of *really* reproducible research [3, 5, 10]. But even if researchers adopted this principle, the use of standard evaluation sets is often not possible, since the evaluation images are often not made available.

Of the three mentioned problems, viz. under specification of algorithms, evaluation sets not available, and different or simply unclear evaluation methods, the latter can be readily solved by proposing standard evaluation methods. The main goal of this paper is thus to present a simple LPL algorithm evaluation method and to use it to evaluate a few relevant LPL algorithms described in the literature over an evaluation set of images obtained under realistic circumstances.

All the software used in this paper is available in Matlab at http://www.livingdata.pt/projects/SCRUARM/publicacoes/.

2. EVALUATION METHODOLOGY

According to Triers and Jain [12], an effective way of measuring the performance of an algorithm is to insert it into the system, and test its overall performance. In this way, the obtained results will reflect the strengths and weaknesses of the algorithm with respect to its operating environment. In the case of LPR systems, this means that a good evaluation method for LPL algorithms would be to use the overall plate recognition success rate. This approach is appropriate whenever the operating environment of the algorithm is fixed.

If the operating environment of the algorithm is not fixed, evaluating its performance by inserting it into a test system using standard algorithms, not necessarily optimized, may have the adverse effect of leading to algorithms which are tuned for the standard environment, though not good in general. To avoid this problem, the evaluation may be performed for all possible combinations of algorithms for the other blocks of the system, but this is often not practical.

A possible approach to these problems would be to start by optimizing the algorithms for the downstream blocks. This would require manually obtaining their input data. In the case of a LPR system, where the downstream block is OCR, it would mean manually locating and segmenting the plate in each of the input test images, so that the located plate image might be fed into the OCR block.

Applying such a scheme to all blocks of a system would require manually obtaining the data transferred between all of its main blocks. While this may seem infeasible at first, in the case of LPR the work involved is perfectly manageable. On the other hand, the obtained data may be used both as input to blocks whose algorithms are being optimized, and as a reference against which to check the actual results of the blocks upstream, which may thus be evaluated separately. This paper proposes an evaluation method for LPL algorithms which uses this approach.

3. EVALUATION METHOD

3.1. Assumptions

The main assumption of the developed evaluation method is that there is always a single relevant plate visible in each image. Another assumption is that the result of applying an LPL algorithm to an image is a set of location hypotheses consisting of a sequence of four plate corners, given in clockwise order and starting in the upper left corner. Since it is assumed that the plates to locate are planar and rectangular, the location of its corners is sufficient to determine which pixels belong to the located plate image and which do not.

3.2. Manually obtaining ground-truth data

Analyzing images manually is generally a daunting task. In the case of LPR, however, the data to be obtained has characteristics which greatly simplify this task: (i) plates are planar and rectangular, (ii) plates have a fixed number of possible models, (iii) their relevant parts consist of characters against a uniform background, (iv) the characters are letters and digits, (v) the relevant characters are all of the same height, though of varying width, and (vi) the characters are horizontally aligned in a small number of rows within the plate.

For measuring LPL accuracy it is not sufficient to know the actual plate limits. It is also very important to know where the characters which compose it are. Hence, the required ground truth data consists of the location of plate and the location of the plate's characters.

3.3. Performance criteria

The LPL evaluation method is based on several performance criteria:

At least one hypothesis The location fails if no LPL hypotheses are generated.

Small number of location hypotheses The number of LPL hypotheses must be small, so that the LPL problem is not transferred entirely to downstream blocks of the overall system.

Plate must not be missed The plate is considered to be missed if all hypotheses miss all plate characters.

Characters must not be missed There is at least one hypothesis for which there are no missing plate characters.

Located region may not be too large It is not enough for a location hypothesis to envelop the characters of the plate in the image: it is fundamental that it is as tight as possible.

The evaluation method in [7] uses a criterion for acceptance of a LPL hypothesis which is based simply on the Jaccard similarity between the bounding box of the real plate and the region produced by the LPL algorithm. Hence, it does not take into account the fact that plate characters are the most important information. This is an important difference relative to the evaluation method proposed here, which states that missing plate pixels are relevant only if they belong to the bounding box of one of the characters.

In this paper, a quality factor ranging from 0 to 1 is calculated for each LPL hypothesis and a global quality factor is calculated for the set of all LPL hypotheses produced for a single image. However, the decision of whether or not to accept as good a given LPL result is not based in this global quality factor.

The global quality factor Q of a given LPL result is

$$Q = \begin{cases} 0 & n_h = 0, \\ q(n_h, n_{h_{\max}}) \times \max\{Q_i, i = 1, \ldots, n_h\} & n_h \neq 0, \end{cases} \quad (1)$$

where n_h is the total number of location hypotheses, $n_{h_{\max}}$ is the maximum number of hypotheses (10 in this work), Q_i is the quality of hypothesis i, and

$$q(E, E_{\max}) = \frac{1}{1 + \dfrac{e^{k\frac{E}{E_{\max}}} - 1}{e^k - 1}}, \quad (2)$$

is a quality function where k is a steepness control parameter. The value of k used was 2.5. The LPL is rejected if $n_h > n_{h_{\max}}$.

The quality of each hypothesis is the geometric average of two qualities, related with the criteria that the LPL result may not be too large and may not miss characters by too much. It is calculated as

$$Q_i = \sqrt{Q_{s_i} Q_{m_i}}, \quad (3)$$

where Q_{s_i} is the quality associated with spurious regions which were deemed to belong to the plate but with actually do not, and where Q_{m_i} is the quality associated with the missing areas which were deemed not to belong to any of the plate character bounding boxes, but which actually do belong to one of these bounding boxes. The quality associated with spurious regions is obtained applying the quality function in (2) to the spurious regions error, which will be soon be introduced, i.e.,

$$Q_{s_i} = q(E_{s_i}, E_{s_{\max}}),$$

where $E_{s_{\max}}$ is the maximum spurious regions error. Hypothesis i is considered to be too large if $E_{s_i} > E_{s_{\max}}$. The quality associated with the missing character regions is calculated for each character, again using equation (2), and geometrically averaged over all characters, i.e.,

$$Q_{m_i} = \sqrt[n_c]{\prod_{j=1}^{n_c} q(E_{m_{ij}}, E_{m_{\max}})},$$

where n_c is the total number of characters in the plate, $E_{m_{ij}}$ is the error associated with missing regions for character j relative to hypothesis i, and $E_{m_{\max}}$ is the maximum missing regions error. Character j is considered to be missed by hypothesis i if $E_{m_{ij}} > E_{m_{\max}}$.

3.4. Location errors

In order for the LPL evaluation to be independent of the camera characteristics and of the actual position and orientation of the imaged plate relative to the camera, the perspective projection is compensated for, so that the criteria are applied to normalized, rectangular plate images.

Since LPL can be seen as the process of segmenting the image into two parts (the plate itself and the rest of the image), the evaluation of location hypothesis errors, viz. the errors related to spurious plate pixels and to missing character pixels, can be seen as a specific case of the more general problem of segmentation evaluation [12, 13]. This issue is an important one in image analysis, though not in the scope of this paper. A good overview of segmentation evaluation techniques can be found in [4].

When assessing whether the location hypothesis is too large, it seems reasonable take into account not only the area of the spurious regions, but also its distance to the real plate location. A good solution seems to be to sum the weights of the spurious region pixels, where the weight of each pixel is proportional to its distance to the real plate location. A similar reasoning may be used for assessing whether characters are missed by the location hypothesis. The measurement used should take into account the missed character pixels, but weighting them according to their distance to the boundary of the character bounding box. These measurements correspond to the two terms of the "distance-weighted shape fidelity" definition in [4].

The evaluation method in [7] does not distinguish between missing and spurious pixels: a single measurement, viz. the Jaccard similarity, is used to measure the quality of the LPL results. It also does not take into account the distance of the pixels in error to their correct location.

The error associated with spurious plate pixels is

$$E_{s_i} = \sum_{p \in R_{p_i} \setminus R_p} w_s\left(d\left(p, R_p\right)\right),$$

where R_p is the set of pixels corresponding to the real plate image, R_{p_i} is the set of pixels corresponding to the plate location hypothesis *i*, $d\left(p, R_p\right)$ is the Euclidean distance between pixel *p* and region R_p, and $w_s(\cdot)$ is a weighting function for the distances, which in this case is the identity function, i.e., $w_s(d) = d$ for all values of *d*.

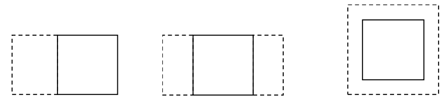
It is usually easier for the downstream blocks to horizontally segment the characters, especially when the plate is slanted, and thus to discard spurious horizontal pixels, than to vertically segment the characters. Hence, it was considered preferable to use a weighting of the distances which would allow larger errors to occur in the direction of largest plate dimension (which usually is the width). A simple way to accomplish this different horizontal and vertical weighting was to normalize the plates into square regions of $\sigma_p n_p \times \sigma_p n_p$ pixels, where n_p was set to 100 and σ_p is a scaling factor which allows the exact size of the square region to be adjusted.

The value of $E_{s_{\max}}$ was empirically set to 250,000 when $\sigma_p = 1$. Figure 1 shows the maximum canonical errors for several values of $E_{s_{\max}}$.

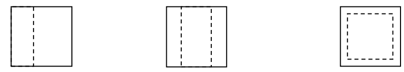
The error associated with missing character pixels is

$$E_{m_{ij}} = \sum_{p \in R_{c_j} \setminus R_{pi}} wm\left(d\left(p, \overline{R_{c_j}}\right)\right),$$

where R_{c_j} is the set of pixels corresponding to the bounding box of character *j* in the real plate image, R_{p_i} is the set of pixels corresponding to the plate location hypothesis *i*, $d\left(p, \overline{R_{c_j}}\right)$ is the Euclidean distance between pixel *p* and the complement of region R_{c_j}, and $w_m(\cdot)$ is a weighting function for the distances, which in this case is the identity function, i.e., $w_m(d) = d$ for all values of *d*. As in the case of spurious plate pixels, the real character bounding boxes are normalized to a square image of $\sigma_c n_c \times \sigma_c n_c$ pixels, where n_c was set to 100



$E_{s_{\max}}$			
125,000	49.5%	34.9%	21.6%
250,000	70.2%	49.5%	29.5%
500,000	99.5%	70.2%	39.9%



$E_{m_{\max}}$			
10,000	13.7%	9.5%	6.9%
20,000	19.5%	13.7%	10.3%
40,000	27.8%	19.5%	15.3%

Figure 1 – Maximum error, for different values of $E_{s_{\max}}$ ($E_{m_{\max}}$), and for different types of error, measured in percentage of the plate (character) dimension in the error direction.

and σ_c is a scaling factor which allows the exact size of the square region to be adjusted.

3.5. Reporting on LPL hypotheses results

The evaluation of a LPL result will result in one of the several classes. If the number of hypotheses is zero, the class is "no location"; if the number of hypotheses is too large, the class is "too many hypotheses"; if all location hypotheses missed the plate, the class is "plate missed"; if all location hypotheses missed at least one character, the class is "characters missed"; otherwise there is at least one location hypothesis which did not miss any characters; if all these hypotheses are too large, the class is "too large"; otherwise the class is "correct".

3.6. Evaluation set

A large collection of 1968 gray-level images acquired at a toll gate has been made available by Brisa S.A. Of the total of 1968 images available, a smaller set of 1422 was chosen by removing images with no visible plate, with some plate characters invisible or illegible, with foreign plates, and with more than one plate. Of the total of 1422 images available, 673 were randomly chosen to be part of the evaluation set.

The Portuguese plates considered in this paper have three possible models: model 1 (97.9% of the evaluation set), model 2 (1,5%) and model 4 (0,6%). The plate bounding box dimensions in the images do not vary much, having an average width of approximately 83.3 pixels (for model 1 plates).

Please send an email to scruarm@brisa.pt to know under which circumstances access to the evaluation images is granted.

4. EVALUATED LPL ALGORITHMS

Some of the LPL algorithms described in the literature were implemented and tested, namely Setchell's [11], Naito *et al.*'s [9], Martín *et al.*'s [8], and Brandão *et al.*'s [2]. The implementation of the algorithms (with the exception of Brandão *et al.*'s) was hindered by the fact that descriptions of the algorithms were not given in sufficient detail.

The evaluation methodology used allows the use of a single parameter to be passed into the LPL algorithms: the expected plate width. The value used was 85 pixels for all algorithms.

5. RESULTS

The results are summarized in Table 1. It must be taken into account that the results correspond to our own implementations of the other author's algorithms, whose descriptions are insufficiently detailed and for which no source code was found.

It can be seen that the best correct location rate is attained by Brandão *et al.*'s algorithm. However, the performance of this algorithm seems to be attained at the cost of a higher processing time.

Brandão *et al.*'s algorithm produces on average a smaller number of hypotheses than its direct competitors. This is due to its multi-stage nature, where oversized hypotheses are sent to downstream stages for refinement, which allows to overall algorithm to be more selective in its generation of location hypotheses. The smallest average number of hypotheses is attained by Naito *et al.*'s algorithm, which also has a modest correct location ratio of 89%.

The average global quality attained by Brandão *et al.* is better than the other algorithms, since it also has a higher correct location ratio. However, its average quality of the first correct location is smaller than the one attained by the other algorithms. Setchell's method, for instance, has an excellent quality of location for the locations which are considered correct. This measure is important, since the better the locations produced, the easier it will be for the downstream blocks in the system to segment and recognize the characters.

Measurement	Setchell	Naito	Martín	Brandão
Correct	93.16	89.00	93.91	97.18
Too large	0.59	3.86	1.78	0.89
Characters missed	3.57	5.79	2.67	1.49
Plate missed	0.59	0.89	1.49	0.15
No locations	2.08	0.30	0.15	0.30
Many hypotheses	0.00	0.15	0.00	0.00
Average hyp.	2.725	1.283	2.819	1.788
Maximum hyp.	9	15	7	7
Global quality	0.871	0.891	0.863	0.925
Quality of first correct location	0.991	0.984	0.981	0.974
Minimum time	0.234	0.296	0.750	1.281
Average time	0.605	0.481	1.157	1.897
Maximum time	1.657	2.532	1.906	10.891

Table 1 – Results attained by tested algorithms (rates in percentage, times in seconds)

6. CONCLUSIONS

An LPL evaluation method has been proposed and made available in the form of a Matlab package. This will allow a fair comparison of different LPL algorithms. An evaluation set of images has also been made available, which together with the evaluation method proposed can be used to establish an LPL algorithm benchmark. Even though the evaluation method proposed already permits interesting and reproducible comparisons between LPL algorithms, it may still be improved in a number of directions.

The authors would like to acknowledge the support of Brisa, ISCTE, and Living Data.

8. REFERENCES

[1] G. Adorni, S. Cagnoni, M. Gori, and M. Mordonini, "Access control system with neuro-fuzzy supervision," *Proc. IEEE Intell. Transp. Systems*, pp. 472-477, 2001.

[2] T. Brandão, M. Menezes de Sequeira, and M. Albuquerque, "Multistage morphology-based license-plate location algorithm," Accepted for publication in the Proceedings of WIAMIS 2004.

[3] J. B. Buckheit and D. L. Donoho, "WaveLab and reproducible research," *Wavelets and Statistics*, vol. 103, pp. 55-81, Springer-Verlag, Berlin, 1995.

[4] P. Correia, *Video Analysis for Object-Based Coding and Description*, PhD thesis, IST, 2002.

[5] J. de Leeuw, "Reproducible research: The bottom line," Statistics Electronic Publication 301, UCLA, 2001.

[6] C. de Mello and R. Lins, "A comparative study on OCR tools," *Proc. Vis. Interface'99*, pp. 224-231, 1999.

[7] K. I. Kim, K. Jung, and J. H. Kim, "Color texture-based object detection: An application to license plate localization," *Proc. 1st Int. Workshop on Pattern Recognition with Support Vector Machines*, vol. 2388, pp. 293-309, 2002.

[8] F. Martín, M. García, and J. L. Alba, "New methods for automatic reading of VLP's (vehicle license plates)," *Proc. Int. Conf. on Sign. Proc., Pattern Recogn., and Applications*, 2002.

[9] T. Naito, T. Tsukada, K. Yamada, K. Kozuka, and S. Yamamoto, "Robust license-plate recognition method for passing vehicles under outside environment," *IEEE Trans. Vehicular Technology*, 49(6), pp. 2309-2319, 2000.

[10] M. Schwab, M. Karrenbach, and J. Claerbout, "Making scientific computations reproducible," *Comp. in Science & Engineering*, 2(6), pp. 61-67, 2000.

[11] C. Setchell, "Applications of Computer Vision to Road-Traffic Monitoring," PhD thesis, University of Bristol, 1997.

[12] O. D. Trier and A. K. Jain, "Goal-directed evaluation of binarization methods," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 17(12), pp. 1191-1201, 1995.

[13] Y. J. Zhang, "A survey on evaluation methods for image segmentation," *Patt. Recognition*, 29(8), pp. 1335-1346, 1996.