Data-Driven Vehicle Identification by Image Matching

Jose A. Rodriguez-Serrano[1], Harsimrat Sandhawalia[1], Raja Bala[2], Florent Perronnin[1], and Craig Saunders[1]

[1] Xerox Research Centre Europe, Meylan, France
[2] Xerox Research Center Webster, NY, USA

Abstract. Vehicle identification from images has been predominantly addressed through automatic license plate recognition (ALPR) techniques which detect and recognize the characters in the plate region of the image. We move away from traditional ALPR techniques and advocate for a data-driven approach for vehicle identification. Here, given a plate image region, the idea is to search for a near-duplicate image in an annotated database; if found, the identity of the near-duplicate is transferred to the input region. Although this approach could be perceived as impractical, we actually demonstrate that it is feasible with state-of-the-art image representations, and that it presents some advantages in terms of speed, and time-to-deploy. To overcome the issue of identifying previously unseen identities, we propose an image simulation approach where photo-realistic images of license plates are generated for desired plate numbers. We demonstrate that there is no perceivable performance difference between using synthetic and real plates. We also improve the matching accuracy using similarity learning, which is in the spirit of domain adaptation.

1 Introduction

This article focuses on vehicle identification from images [1–4], which finds a variety of applications in the transportation industry. For instance, on-board cameras in police cars and buses are employed to identify vehicles incurring traffic violations; fixed cameras are installed as part of the traffic infrastructure to automate payments in tolls or parkings, and cameras in portable devices and even mobile phones allow identifying vehicles by parking enforcement staff or for vehicle social networks[1].

A vast number of off-the-shelf automatic license plate recognition (ALPR) [3, 4, 1] tools exist which detect and recognize plate characters in vehicle images. Despite the perception that ALPR is a solved problem, the identification of vehicles with low-quality cameras (such as from mobile phones), at high speed, or in countries with a vast number of plate designs (such as the USA) still poses research challenges. Also, industry experts estimate that the time to deploy an

[1] See, for instance, `www.bump.com`

A. Fusiello et al. (Eds.): ECCV 2012 Ws/Demos, Part II, LNCS 7584, pp. 536–545, 2012.

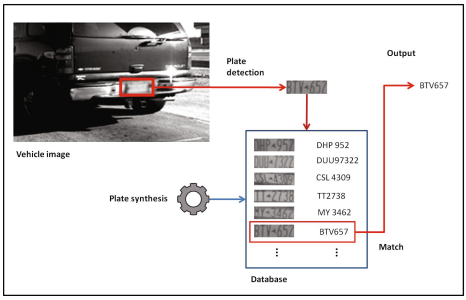

Fig. 1. Overview of the system. *Note: For privacy reasons, license plate images and numbers are fictitious.*

existing ALPR system in a new state or country, including sample acquisition, manual annotation, re-training and fine-tuning can take months.

In this paper, we move away from traditional ALPR techniques and advocate for data-driven approaches for vehicle identification. Here, given a plate image region, the idea is to search for a near-duplicate image in an annotated database; if found, the identity of the near-duplicate is transferred to the input region. Our goal is to show that with state-of-the-art image representations for retrieval, such as Fisher vectors [5, 6], the data-driven approach is feasible in terms of speed and robustness, and presents advantages. First, *a very significant and somewhat counter-intuitive finding of this paper is that a matching approach can actually outperform ALPR in close-world systems* - i.e. when for all the vehicles to identify an annotated instance exists in the database. Secondly, the time-to-deploy is reduced: image matching requires plate-level annotations while ALPR demands character-level segmentations and definition of character sets or language models, which is more costly to obtain. And thirdly, the accuracy of image matching degrades less critically than for ALPR when moving to low-quality cameras.

However, an obvious issue of near-duplicate matching is that a vehicle which has not been previously registered in the database cannot be identified. For example, an authority needs to perform a search for a vehicle with known license plate number but the image database might not contain that plate. To overcome this difficulty, we propose to use a photo-realistic rendering approach [7] to generate a license plate image for "wanted" plates and add them to the database. We demonstrate that using those simulated images does not impact the matching accuracy, and also that this accuracy can be boosted using a better similarity function obtained through *similarity learning*, in the spirit of recent works on domain adaptation [8].

An overview of the method is shown in Fig. 1. The rest of the article is structured as follows. §2 describes the matching approach and image descriptor. §3 elaborates on the image simulation approach. §4 reports the experimental validation. In §5 the conclusions are drawn.

2 Data-Driven Vehicle Identification

So far the predominant technique for image-based vehicle identification has been ALPR. We refer the reader to the survey [3] and the more specific papers [1, 4, 2] for details. However, in recent years there has been an impressive progress in image retrieval, both in terms of discriminative power of features and compression/indexing, and state-of-the-art methods are capable of comparing an image with a database of millions of images in milliseconds [9] with high accuracy and modest computing resources.

Inspired by these recent results, this article proposes a data-driven view for vehicle identification. Here, given an image of an unknown vehicle, captured either from the infrastructure, from an on-board camera or from a mobile device, the approach we propose is to describe the license plate sub-image and match it against a database of annotated license plate images with state-of-the-art retrieval techniques. If a near-duplicate image is found, the annotated license plate number is transferred to the input image. Similar ideas have been used with success to solve the problems of image geo-tagging [10] or image completion [11] , among others, but we are not aware of previous attempts for ALPR (or in general any scene text recognition).

This section discusses the plate image descriptor and matching approaches. We omit the details on plate localization here as this is not the core of the work. For reference, Fig. 4 illustrates the achieved level of plate segmentation.

Plate Features. For the license plate image descriptor we extract Fisher vectors [5] since these have been reported to be state-of-the-art descriptors for image categorization tasks [6] and very large-scale image retrieval [9], and show robustness in the range of photometric and geometric variability present in our application.

In a nutshell, Fisher vectors work by aggregating local patch descriptors into a fixed-length representation. First, SIFT patches are extracted at multiple scales on a regular grid, and their dimensionality is reduced using principal component analysis (PCA). A visual vocabulary is built by estimating a Gaussian mixture model (GMM) with patch descriptors extracted from a held-out set of images. The Fisher vector is computed as the derivative of the log-likelihood with respect to the GMM parameters. For instance, if we consider the means only, it can be shown that the expression is given by

$$f_{id} = \gamma_i(\mathbf{x}_t) \left[\frac{x_{t,d} - m_{i,d}}{(S_{i,d})^2} \right], \tag{1}$$

where $\gamma_i(\mathbf{x}_t)$ is the soft-assignment probability of the tth patch to the ith Gaussian, $x_{t,d}$ is the dth component of the ith patch, and $m_{i,d}$ and $S_{i,d}$ are the dth components of the mean and standard deviations of the ith Gaussian, assuming diagonal covariances. Here, $i = 1 \dots K$ and $d = 1 \dots D$. If we only use the derivatives with respect to the mean[2], then the resulting Fisher vector is a

[2] We discard derivatives with respect to the weights or the variance since they only bring a small improvement but do impact computation time and signature sizes.

concatenation of the $K \times D$ elements f_{id}. We also apply the square-rooting and ℓ_2-normalization of [5], and make use of spatial pyramids to account for spatial information.

Matching. Since the Fisher vector is an explicit embedding of the Fisher kernel, the corresponding similarity measure between two such image descriptors $\mathbf{x}$ and $\mathbf{y}$ is the dot product $\mathbf{x}^T\mathbf{y}$. A candidate plate is compared against all images in a database and we assign the identity of the closest match, provided the similarity is sufficiently high.

The advantage of matching with respect to ALPR systems is that character segmentation is not necessary, and that it is "agnostic" of the character set, layout and design of the plates; it just assigns an identity by finding a near-duplicate image. This is advantageous e.g. for US plates, which present stacked characters, graphical symbols or complex backgrounds for which character recognition techniques have difficulties dealing with. An illustration of these difficulties is shown in Fig. 2.

Fig. 2. License plate templates showing typical difficulties for ALPR systems: state symbols, stacked characters, complex backgrounds

Also, the time to deploy a matching platform in a new location is much shorter than for an ALPR system since the latter requires manual character-level annotation and learning the character classifiers.

3 License Plate Image Simulation and Matching

A requirement of data-driven identification is that each potential identity needs to be represented at least by one example in the database. This requirement is fulfilled in "close-world" applications such as entry-exit matching, identification of returning vehicles, or access control of pre-registered vehicles. However, in many situations it is also useful to search for or identify unexpected or previously unseen vehicles.

We propose to bypass this limitation by *simulating* images of license plates for previously unseen vehicles, following the approach of [7]. This was used to generate a dataset of realistic license plate images for character classifiers for ALPR in an inexpensive way. Nevertheless, to the best of our knowledge, this method has not been applied before to produce images to populate a dataset for matching[3].

[3] In computer vision, data simulation has been used to synthesize training data for pedestrian detectors [12], 3D object recognition [13] or scene text classification [14]. In the document analysis literature we find works which synthesize word images for search [15, 16] but the complexity of these document images is much smaller than for outdoor license plates.

3.1 Image Synthesis and Photo-Realistic Transformation

License plate simulation consists of two main steps: (i) synthesizing an ideal license plate image, and (ii) applying transformations to reproduce the characteristics of real license plate images. Next we summarize the process. We refer the reader to [7] for the details.

Synthesizing an Ideal License Plate Image. Synthesizing an ideal license plate image basically requires defining the background template, the font properties, and the valid character sequences and positions. This information can be obtained from existing specifications. Also, existing resources such as on-line repositories of license plate templates[4], character erasure methods or software tools that produce TrueType fonts from character examples can be exploited to assist this one-off procedure.

Thus, license plate images can be produced by overlaying the desired character sequences with the appropriate font in the template at adequate positions. See first steps of Fig. 3(a) for illustration.

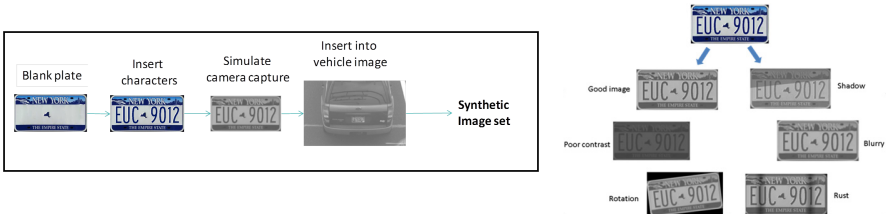

Fig. 3. (a) Overview of license plate simulation. (b) Distortions considered.

Photo-Realistic Transformation. A number of transformations and distortions are applied next to produce realistic images which simulate camera capture. Fig. 3(b) illustrates some operations considered.

For instance, in ALPR it is common to use near-infrared cameras since they present better signal-to-noise ratio across different illuminations and day/night variations. Thus, a color-to-infrared transform is needed. Here, a transformation $I_{IR} = \max(w_R R, w_G G, w_B B)$ is applied to (R, G, B) pixel intensities, where w_R, w_G, and w_B are relative material transmittances in the R/G/B channels and are adjusted from RGB-IR value pairs.

Then, image brightness and contrast are adjusted using a simple affine transformation $I_{out} = \alpha I_{IR} + \beta$ to the IR intensities, the parameters also being learned from contrast value pairs of real and synthetic plates.

A number of other transformations are considered: blur (in the form of a resolution conversion which was found to be more effective than spatial filtering), or adding shadows, rust or spurious noise.

[4] e.g. `www.blankplates.com`

Finally, to ensure that we produce tight plate regions in similar conditions as for the real plates (see section 2), simulated license plate images are inserted into a set of full-vehicle images with the four license plate corners manually marked. The corresponding affine transformation between the simulated image plane and the 4 corners is determined. After that, the plate localization algorithm is run to yield the tight plate regions. An example of a real and simulated tight regions as obtained by the detection algorithm is shown in Fig. 4.

Fig. 4. Plate regions of simulated (left) and real license plates (right)

3.2 Similarity Learning

To increase the efficiency of matching synthetic and real plate images, we employ similarity learning. A number of distance and similarity learning methods have been proposed in the machine learning community (see, e.g. [17, 18]) which basically seek a projection of the data more suitable for nearest-neighbor classification or ranking. Interestingly, asymmetric distance learning has been employed in the computer vision literature to match images of two different domains [8], which is close to our problem.

Denote $\mathbf{s}$ the descriptor of a simulated plate image and $\mathbf{r}$ that of a real plate image. We search for a function of the form $k(\mathbf{r}, \mathbf{s}) = \mathbf{r}^T \mathbf{W} \mathbf{s}$. In a default setting where $\mathbf{W}$ equals the $L \times L$ identity matrix, k is the dot product which is the standard measure of similarity between Fisher vectors. However, we aim at finding a more suitable $\mathbf{W}$ inspired by the large-margin supervised semantic indexing [17] approach.

We search for a matrix $\mathbf{W}$ which minimizes the following loss

$$\sum_{(\mathbf{r}, \mathbf{s}^+, \mathbf{s}^-)} \max\{0, 1 - k(\mathbf{r}, \mathbf{s}^+) + k(\mathbf{r}, \mathbf{s}^-)\}. \tag{2}$$

where $\mathbf{s}^+$ is a simulated sample with the same identity as the real sample $\mathbf{r}$, and $\mathbf{s}^-$ has a different identity. Note that minimizing the loss encourages $k(\mathbf{r}, \mathbf{s}^+) > k(\mathbf{r}, \mathbf{s}^-) + 1$, i.e. pairs with the same identity should have a higher similarity than non-matching pairs, which is a desirable property of a similarity function, and the $+1$ acts as a "safety" margin to promote generalization.

This loss function can be optimized using Stochastic Gradient Descent (SGD) [19]. Following straightforward derivations, it is possible to show that the training procedure consists in repeating the two following steps: (i) sample a triplet $(\mathbf{r}, \mathbf{s}^+, \mathbf{s}^-)$ randomly, and (ii) perform the gradient update

$$W \leftarrow W + \eta \mathbf{r}(\mathbf{s}^+ - \mathbf{s}^-)^T \tag{3}$$

if the loss $\max\{0, 1 - k(\mathbf{r}, \mathbf{s}^+) + k(\mathbf{r}, \mathbf{s}^-)\}$ is positive, where η is a learning rate.

We initialize $\mathbf{W}$ with the identity matrix so we start from a reasonable solution (dot product).

4 Experiments

The proposed data-driven vehicle identification method is validated using real data collected in open-road tolls in several installations in the USA. Vehicles driving at about 50mph are detected by loop sensors which trigger an image capture of the full vehicle. Images contain a vast variety of vehicle types and times of day and the different locations imply different camera perspectives and from front/back directions with expensive tolling cameras. The images come in resolutions of 768×484 and 1920×512 and contain near-infrared intensity values (as this has a better signal-to-noise ratio for the license plate region across a range of illuminations, which for instance permits capturing license plates even at nighttime). The equipment and configuration is optimized for ALPR. An example image is shown in Fig. 1.

Note that this corresponds to the case where vehicles are identified by the infrastructure but the method is general and could apply to other settings such as on-board cameras and mobile phones (note the resolution and high vehicle speed). We obtain a dataset for our experiments by running a license plate detection algorithm.

Image Matching for Vehicle Identification. We evaluate image matching alone as a vehicle identification method and compare it to two industrial ALPR products representing two off-the-shelf baselines for identification. We will demonstrate not only that image matching is feasible for identification, but also that matching can outperform ALPR in some cases.

To that end, we use a "database" of about 35K images (corresponding to one week of data) and an evaluation set of 11K samples, with license plate number ground-truth. For each query image, we determine the most similar image in the database and output the license plate number of the match and the similarity value as confidence score. This is directly comparable to the ALPR baselines which also output a license plate number associated with a confidence.

Figure 5(a) shows the accuracy-reject curves for the two ALPR systems and the image matching, where we use the dot product between Fisher vectors as the similarity measure. Reject represents the fraction of samples with confidence below a threshold, and accuracy the fraction of samples with confidence above the threshold whose outputs match the ground-truth annotations. We observe that the ALPR systems reach a maximum accuracy in the range 80%-90%, for reject rates of about 40%. The reason why the accuracy never reaches 100% is because some plates contain graphical symbols, stacked characters and other artifacts that the ALPR systems are not able to deal with. In contrast, matching yields a much higher maximum accuracy, close to 100%, since it is agnostic on character dictionary, language, etc. However, this accuracy is only reached at a very high reject rate (about 90%) and quickly drops for smaller reject rates. This

is because most of the plates in the query set are not present in the database and are rejected by the system.

However, the situation changes if the query images do have at least one "true" match in the database. If we consider only the set of about 5K queries which have an instance in the database, we obtain the result of Fig. 5(b). Now not only image matching reaches close to 100% at a much smaller reject rates, but it also clearly outperforms the ALPR systems both in accuracy and reject rate.

Thus, *image matching is a good vehicle identification method, actually better than commercial ALPR systems, provided that annotated examples are available.*

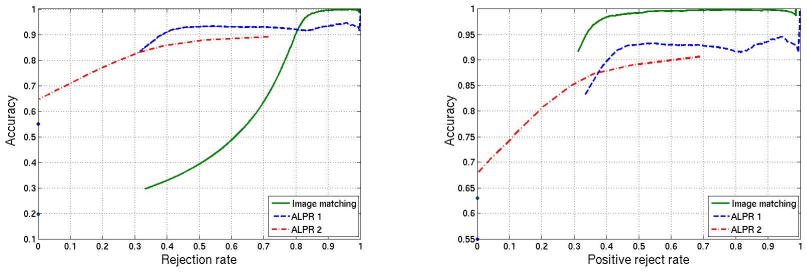

Fig. 5. Accuracy of image matching for identification: (a) when not all queries have a matching instance in the database, (b) when all the queries have an instance in the database

Use of Simulated Plates. In this section we evaluate the capability of the system to identify previously unseen vehicle identities through image simulation. We consider a set of 582 plate identities for which we synthesize a random number of plate images with different distortions using the method described in section 3, resulting in a set of 3,476 images. A set of 582 real images of the same identities is selected, to query with those and assess whether they are found in the database. This corresponds to a situation where there is a list of identities to be checked and a system captures images and compares each image with the synthesized images of the wanted identities, using again the dot product as similarity measure (i.e. no metric learning yet; experiments with metric learning are reported in the next paragraph). The error-reject characteristic of this setting is depicted as a solid, blue curve in Fig. 6 .

Then we repeat the experiment by replacing each synthetic image with a real image of the same license plate number (with random selection when there are more real images than synthetic ones). The corresponding curve is displayed in dashed red in Fig. 6. We observe that the curves corresponding to a synthetic and a real database are essentially on par. Therefore, we conclude that using a synthesized license plate has essentially the same effect as the real plate. *Thus, generating license plate images is an effective way to enable the search for previously unseen vehicle identities.*

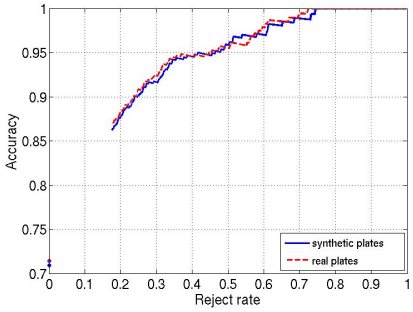

Fig. 6. Comparison between accuracies of real and synthetic plates in a retrieval task

Similarity Learning. We evaluate the similarity learning method of section 3.2 for the task of matching real to synthetic images. We essentially use the same set as for the previous experiment but we divide the query and database into two halves both containing synthetic and real plates for training and evaluation. We report both identification accuracy rates at 0% reject (acc0) and at 20% reject(acc20) and the mean average precision (mAP) to assess whether the learned similarity also performs well in ranking the database images beyond the first-best result. Results are summarized in Table 1.

Table 1. Results with similarity learning

	acc0(%)	acc20(%)	mAP(%)
No learning	78.6	87.8	77.2
Similarity learning	**81.9**	**92.7**	**81.5**

These results are with 100 training epochs and $\eta = 10^{-2}$. Table 1 shows that the metric learning algorithm improves both the accuracy and the mAP.

5 Conclusions

This article demonstrates that a data-driven view of vehicle identification without ALPR is feasible, and actually more accurate than ALPR in some scenarios. We believe that entry/exit control or matching for assisting manual plate verification are two applications that can benefit from the proposed method.

The proposal has been validated on images captured from the infrastructure. Although validation for on-board or mobile device cameras has not been performed, we believe the method is general and also applicable to these scenarios. Actually, in a set of experiments not reported here, it was observed that reducing the resolution of the images does not degrade the image matching accuracy as severely as it did for ALPR.

A line of future work is to go for compressed signatures in applications where the large database size is of concern.

References

1. Arth, C., Limberger, F., Bischof, H.: Real-time license plate recognition on an embedded DSP-platform. In: CVPR (2007)
2. Donoser, M., Arth, C., Bischof, H.: Detecting, Tracking and Recognizing License Plates. In: Yagi, Y., Kang, S.B., Kweon, I.S., Zha, H. (eds.) ACCV 2007, Part II. LNCS, vol. 4844, pp. 447–456. Springer, Heidelberg (2007)
3. Anagnostopoulos, C.N.E., Anagnostopoulos, I.E., Psoroulas, I.D., Loumos, V., Kayafas, E.: License plate recognition from still images and video sequences: A survey. IEEE Trans. on Intelligent Transportation Systems 9 (2008)
4. Chang, S.L., Chen, L.S., Chung, Y.C., Chen, S.W.: Automatic license plate recognition. IEEE Trans. on Intelligent Transportation Systems 5, 42–53 (2004)
5. Perronnin, F., Sánchez, J., Mensink, T.: Improving the Fisher Kernel for Large-Scale Image Classification. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part IV. LNCS, vol. 6314, pp. 143–156. Springer, Heidelberg (2010)
6. Chatfield, K., Lempitsky, V., Vedaldi, A., Zisserman, A.: The devil is in the details: an evaluation of recent feature encoding methods. In: BMVC (2011)
7. Bala, R., Zhao, Y., Burry, A., Kozitsky, V., Fillion, C., Saunders, C., Rodriguez-Serrano, J.A.: Image simulation for automatic license plate recognition. In: Proceedings of SPIE, vol. 8305 (2012)
8. Kulis, B., Saenko, K., Darrell, T.: What you saw is not what you get: Domain adaptation using asymmetric kernel transforms. In: CVPR (2011)
9. Jégou, H., Perronnin, F., Douze, M., Sánchez, J., Pérez, P., Schmid, C.: Aggregating local image descriptors into compact codes. IEEE Trans. on PAMI (2011)
10. Hays, J., Efros, A.A.: im2gps: estimating geographic information from a single image. In: CVPR (2008)
11. Hays, J., Efros, A.A.: Scene completion using millions of photographs. ACM Trans. on Graphics 26 (2007)
12. Marin, J., Vazquez, D., Gerenimo, D., Lopez, A.M.: Learning appearance in virtual scenarios for pedestrian detection. In: CVPR (2010)
13. Schels, J., Liebelt, J., Schertler, K., Lienhart, R.: Synthetically trained multi-view object class and viewpoint detection for advanced image retrieval. In: ICMR (2011)
14. Wang, K., Babenko, B., Belongie, S.: End-to-end scene text recognition. In: ICCV (2011)
15. Konidaris, T., Gatos, B., Ntzios, K., Pratikakis, I., Theodoridis, S., Perantonis, S.J.: Keyword-guided word spotting in historical printed documents using synthetic data and user feedback. IJDAR 9, 167–177 (2007)
16. Rodríguez-Serrano, J.A., Perronnin, F.: Synthesizing queries for handwritten word image retrieval. Pattern Recognition 45, 3270–3276 (2012)
17. Bai, B., Weston, J., Grangier, D., Collobert, R., Chapelle, O., Weinberger, K.: Supervised semantic indexing. In: CIKM (2009)
18. Weinberger, K., Saul, L.: Distance metric learning for large margin nearest neighbor classification. JMLR (2009)
19. Bottou, L.: Stochastic Learning. In: Bousquet, O., von Luxburg, U., Rätsch, G. (eds.) Machine Learning 2003. LNCS (LNAI), vol. 3176, pp. 146–168. Springer, Heidelberg (2004)