

Classifier training based on synthetically generated samples

Hélène Hoessler^{1,2}, Christian Wöhler¹, Frank Lindner¹, and Ulrich Kreßel¹

¹DaimlerChrysler AG, Group Research, Environment Perception
P. O. Box 2360, D-89013 Ulm, Germany

²Ecole Nationale Supérieure de Physique de Strasbourg
BP 10416, F-67412 Illkirch CEDEX, France

Abstract. In most image classification systems, the amount and quality of the training samples used to represent the different pattern classes are important factors governing the recognition performance. Hence, it is usually necessary to acquire a representative set of training samples by acquisition of data in real-world environments. Such procedures may require considerable efforts and furthermore often generate a training set which is unbalanced with respect to the number of available samples per class. In this contribution we regard classification tasks for which each real-world training sample is derived from an ideal class representative which undergoes a geometric and photometric transformation. This transformation depends on system-specific influencing quantities of the image formation process such as illumination, characteristics of the sensor and optical system, or camera motion. The parameters of the transformation model are learned from object classes for which a large number of real-world samples are available. For each individual real-world sample a set of model parameters is derived by correspondingly fitting the transformed ideal sample to the observed sample. The obtained probability distribution of model parameters is used to generate synthetic sample sets for all regarded pattern classes. This training approach is applied to a vehicle-based vision system for traffic sign recognition. Our experimental evaluation on a large set of real-world test data demonstrates that the classification rates obtained for classifiers trained with synthetic samples are comparable to those obtained based on real-world training data.

1 Introduction

In many image classification systems, the size and quality of the training samples used to represent the different training classes are important factors governing the recognition performance. Hence, usually a representative set of training samples has to be acquired by acquisition of data in real-world environments. This procedure often requires considerable efforts but does at the same time not guarantee that each pattern class is sufficiently well represented to obtain a classification system of high generalisation performance.

In many real-world classification tasks, models of the objects to be recognised are available. Such models can be used to synthetically generate training

data for classifiers. In the domain of document image recognition, models of image defects have been studied in some detail. The recognition performance of document recognition algorithms strongly decreases when the image quality is degraded e. g. by printing or scanning. In [1, 2] a parametric document degradation model is proposed based on the physics of printing and imaging. The local deformations of the characters are used to determine a global transformation for the complete document. The parameters of the degradation model are chosen to fit a population of real image documents. The model is then used to generate synthetic data for the training of classifiers, where the synthetic data are inferred from an “ideal” image of the page subjected to the degradation model.

In the domain of industrial machine vision, virtual images obtained by means of a raytracing software from a CAD model are utilised to generate samples to train a classifier for the recognition of holes in industrial parts [7]. This method allows the modification of spatial position, illumination, surface reflectance properties, and color. Reasonable recognition rates are obtained on real test data.

In [3] a method is presented for improving the performance of correlation-based template matching systems. An existing set of real training shapes is extended by virtual shapes in order to improve representational capability. An integrated clustering and registration approach partitions the original shape samples into clusters of similar and registered shapes. Based on the determined cluster parameters, realistic virtual shape samples can be generated. The method is applied to pedestrian detection by correlation matching based on distance transforms.

In this contribution we will regard classification problems in which a real-world training sample is derived from an ideal class representative which undergoes a geometric and photometric transformation due to the image formation process. For each individual real-world sample a set of model parameters is derived by correspondingly fitting the transformed ideal sample to the observed sample. Based on one or several object classes for which a large number of samples is available, probability distributions for the model parameters are obtained, which are utilised to generate synthetic sample sets for all pattern classes. This training approach will be applied to the task of traffic sign recognition.

2 Parametric modelling of real-world training data

The basic concept of the described method for synthetically generating training samples is to consider that an observed sample M_{obs} is an ideal prototype M_{ideal} subjected to a parametric transformation $T(\phi)$ such that

$$M_{\text{obs}} = T(\phi)M_{\text{ideal}}, \quad (1)$$

where the vector ϕ contains the transformation parameters. The model $T(\phi)$ may include geometric distortions (e. g. affine or projective) due to variable viewing direction and perspective, photometric effects such as shading or specular reflections, motion blur due to a non-stationary camera, image imperfections caused by the point spread function of the optical system, or greylevel anomalies due

to nonlinear image sensors. A parameter vector ϕ of the transformation is determined for each individual real-world training sample by minimising the mean squared greylevel difference $E(\tilde{\phi})$ between the observed pattern M_{obs} and the transformed ideal prototype $T(\tilde{\phi})M_{\text{ideal}}$ according to

$$\phi = \arg \min_{\tilde{\phi}} E(\tilde{\phi}) \quad \text{with} \quad E(\tilde{\phi}) = \left\| M_{\text{obs}} - T(\tilde{\phi})M_{\text{ideal}} \right\|^2. \quad (2)$$

We will assume that the probability distribution $P(\phi)$ of the parameter vector ϕ is independent of the pattern class as long as the characteristics of the image acquisition system remain unchanged. Hence, the model parameter distributions are determined based on a small number of well represented pattern classes for which many real-world training samples are available. According to the inferred probability distributions of the transformation parameters, a synthetic training sample M_{synth} is obtained by

$$M_{\text{synth}} = T(\phi)M_{\text{ideal}}, \quad (3)$$

where the parameter vector ϕ is randomly drawn from the previously determined probability distribution $P(\phi)$. An arbitrarily large number of training samples can now be generated readily for each desired pattern class described by its ideal prototype M_{ideal} , respectively.

3 Application to the task of traffic sign recognition

3.1 Outline of the recognition system

We will examine the method described in Section 2 for generating synthetic training samples in the context of a vehicle-based real-time vision system for the recognition of speed limit traffic signs. In this system, a greyscale CCD camera of 640×480 pixels image resolution equipped with a lens of 12 mm focal length, yielding a horizontal viewing angle of about 40 degrees, continuously acquires greyscale video frames of 12 bits pixel depth. The object detection stage consists of a Hough transform [5] extracting circular shapes from each video frame. The detected object hypotheses are tracked across time. Regions of interest (ROIs) are cropped around the object hypotheses, scaled to a uniform size of 19×19 pixels, and normalised to zero mean greylevel and unit variance. A principal component analysis (PCA) is applied to the pre-processed ROIs. It turns out that the loss of information is negligible if only the scores on the 50 most significant principal components are regarded. This reduced set of features is used for classifying the ROI, where a complete quadratic polynomial classifier [9] is employed. The recognition problem corresponds on the one hand to separating the different traffic sign classes from each other, on the other hand to distinguishing between the general ‘‘traffic sign’’ and the ‘‘garbage’’ class that consists of arbitrary ROIs selected by the detection procedure not containing a traffic sign. A track-specific class assignment can be obtained by averaging the single-image class assignments for each track.

3.2 Geometric and photometric transformation model

The transformation model $T(\phi)$ as defined in Section 2 now has to take into account the camera viewpoint, the detection quality, especially the deviation between detected and true centre of a traffic sign, and the illumination conditions. The model fit is applied to the ROIs extracted by the detection stage before scaling and intensity normalisation.

While in principle a projective transformation is necessary to describe the image formation process geometrically, it is sufficient to approximate the projective transformation by an affine transformation due to the relatively small size of the traffic signs. This approximation reduces the number of model parameters and increases the robustness of the model fit. The affine transformation is denoted by $A(t_x, t_y, s_x, s_y, \alpha)$, where the horizontal and vertical translation parameters t_x and t_y describe the accuracy at which the centre of the circular pattern has been detected, while the scale parameters s_x and s_y denote the pattern size. The rotation angle α depends on the relative orientation of the camera with respect to the object. We found empirically that in all our data sets the skew parameter of the affine transformation is negligible and can be set to zero.

Our photometric model is a linear approach described by a gain a and an offset b . This model is appropriate for the linear CCD sensor used in our system but should be replaced by a nonlinear model for CMOS cameras with a logarithmic or linear-logarithmic characteristic curve. More complex models which e. g. account for non-uniform illumination across the object, shading effects or specular reflections may be used instead, if necessary. Furthermore, we found that our real training samples are not perceivably affected by the point spread function (PSF) of the lens, due to the relatively small size to which the samples are normalised (cf. Section 3.3). Hence, it is not necessary here to model the effect of the lens PSF. Possible extensions of the transformation model are discussed in Section 3.4.

According to the described transformation model, a synthetic sample M_{synth} is derived from the class prototype M_{ideal} according to

$$M_{\text{synth}} = a [A(t_x, t_y, s_x, s_y, \alpha)M_{\text{ideal}}] + b. \quad (4)$$

Our method to estimate the parameters of the transformation model is similar to the approach of image registration [4], which involves the optimisation of a quality criterion measuring the correspondence between a target image and an observed image that undergoes a predefined transformation. In our scenario, the target image corresponds to the class prototype M_{ideal} while the observed image M_{obs} is taken from our traffic sign database.

In our system, the quality criterion corresponds to the mean squared greylevel difference between the class prototype and the observed image, such that the model parameters can be obtained according to Eq. (2). We utilise the Levenberg-Marquardt method to determine the set of model parameters and initialise the optimisation algorithm with the result of a grid search across the space of model parameters.

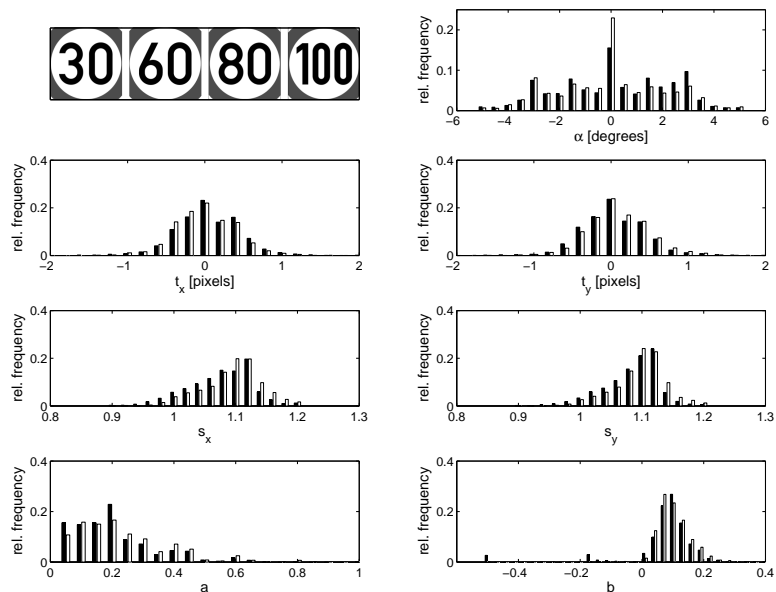


Fig. 1. Ideal class prototypes (upper left) and distributions of model parameters utilised to generate the synthetic training samples (rotation angle α , horizontal and vertical translation t_x and t_y , horizontal and vertical scale s_x and s_y , gain a , and offset b). Filled bars denote parameters inferred from class “60” and empty bars those inferred from class “80”.

3.3 Distributions of model parameters

In order to construct the probability distribution of the transformation parameters we applied the parameter estimation algorithm to a set of 5489 samples of class “60” and 4277 samples of class “80”. These data sets do not overlap with the training and test data regarded later on to determine the classification performance (cf. Section 3.4). The corresponding histograms of the obtained distributions of model parameters are shown for both classes in Fig. 1. The distribution of the rotation angle is centred around zero degrees and obtains typical values between -4 and $+4$ degrees. The horizontal and vertical translations t_x and t_y are centred around zero pixels and typically display absolute values of less than 1 pixel. The distributions of the horizontal and vertical scaling factors s_x and s_y are centred around 1.1 and skewed towards larger values, where a scaling factor of 1 corresponds to a detected diameter of the traffic sign of 19 pixels, the size to which all samples are normalised prior to classifier training (cf. Section 3.4). The gain parameter a covers a continuum of values between 0.05 and 0.4 while the offset b displays a distribution with a narrow peak around

0.1, where these values are valid for pixel intensities normalised to the interval between 0 and 1.

For the two sets of parameter vectors, the covariance matrix reveals that both the scaling parameters s_x and s_y and the photometric parameters a and b are not independent of each other, showing correlation coefficients between 0.4 and 0.5, respectively. All other possible pairs of model parameters display pairwise correlation coefficients of the order 10^{-2} and will thus be regarded as uncorrelated. Hence, the joint probability distribution $P(\phi)$ can be written as a product of several one- and two-dimensional probability distributions:

$$P(\phi) = P(t_x, t_y, \alpha, s_x, s_y, a, b) = P(t_x)P(t_y)P(\alpha)P(s_x, s_y)P(a, b). \quad (5)$$

To examine if the probability distribution of the model parameters is independent of the pattern class, we applied the Kolmogorov-Smirnov test [6] to the two sets of histograms derived for the “60” and the “80” pattern class, respectively. For the joint probabilities $P(s_x, s_y)$ and $P(a, b)$ the test was carried out using Bayes’ theorem in the form $P(s_x, s_y) = P(s_x|s_y)P(s_y)$ and $P(a, b) = P(a|b)P(b)$. Here, the probability distributions $P(s_y)$ and $P(b)$ were inferred directly from the computed set of parameter vectors, while subhistograms describing the conditional probabilities $P(s_x|s_y)$ and $P(a|b)$ were constructed for 20 narrow intervals of s_y and b , respectively. For all model parameters, the Kolmogorov-Smirnov test yields probabilities between 60% and 84% for the hypothesis that the observed probability distributions obtained for the “60” and the “80” pattern class are generated by the same underlying statistical law. This result justifies our initial assumption that the probability distributions of model parameters can be regarded as independent of the pattern class. Consequently, for each model parameter (or pair of model parameters) the respective average probability distribution was used to generate synthetic training samples according to Eq. (3).

3.4 Comparison of classification performance

In this section we will compare the performance of a classifier trained with real traffic sign samples to that of an identical classifier trained with traffic sign samples synthetically generated according to Eq. (3) with model parameter vectors ϕ randomly drawn from the probability distribution $P(\phi)$ determined in Section 3. We regard four traffic sign classes, i. e. “30”, “60”, “80”, and “100”, and a garbage class. The set of garbage patterns consists of real-world samples and is identical in all described experiments. Typical training patterns, both real and synthetically generated, are shown in Fig. 2.

We utilise a complete quadratic polynomial classifier architecture as described in [9]. For a given input pattern, the scores on the 50 most significant principal components of the respective training set serve as an input feature vector to the classifier. The output of the classifier consists of K real-valued numbers d_k with $k = 1, \dots, K$, where K corresponds to the number of classes. In our scenario it is thus $K = 5$. The training labels are chosen such that for a training sample of class l the desired classifier output values \tilde{d}_k are set to $\tilde{d}_k = 1$ for $k = l$ and to $\tilde{d}_k = 0$ otherwise.

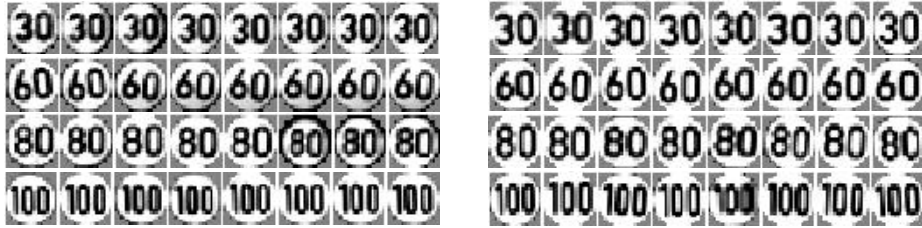


Fig. 2. Real (left) and synthetically generated (right) training samples of the four regarded traffic sign classes.

Traffic sign	Number of real training samples	Number of synthetically generated training samples	Number of real test samples
30	1578	9972	1922
60	6554	9951	9063
80	9264	9952	13673
100	9136	9962	11344
Garbage	14844	14844	23187

Table 1. Composition of the training and the test set utilised for our evaluation of classifier performance.

In the recall phase, the decision to which class l an input pattern is assigned is obtained based on the expression

$$l = \arg \max_k \left(\left\{ \{d_k\}_{k=1, \dots, K-1}, c \cdot d_K \right\} \right). \quad (6)$$

Hence, the maximal classifier output defines the pattern class, where the output denoting the “garbage” class K is scaled by the real-valued factor c . Varying c yields the receiver operating characteristics (ROC) curve of the classifier, depicting the trade-off between the false positive rate, i. e. the fraction of garbage patterns erroneously classified as traffic signs, and the rate of correctly recognised traffic sign samples. The confusion error, denoting the fraction of traffic sign samples assigned to the wrong traffic sign class, is independent of c . All classification errors will be reported on the test set consisting of real samples.

The two training sets utilised in our experiments contain 26532 real and 39837 synthetically generated traffic sign samples, respectively. An identical set of 14844 garbage samples was used in both training sets. The composition of the training and test data is listed in detail in Table 1. The obtained ROC curves are shown in Fig. 3. If we set $c = 1$ in Eq. (6), we obtain false positive rates of 0.14% and 0.08% and rates of correctly recognised traffic sign samples of 98.6% and 97.2% for the classifier trained with real and synthetically generated samples, respectively. The confusion error amounts to 0.13% and 0.003%, respectively. These results show that within the regarded range of very low false positive

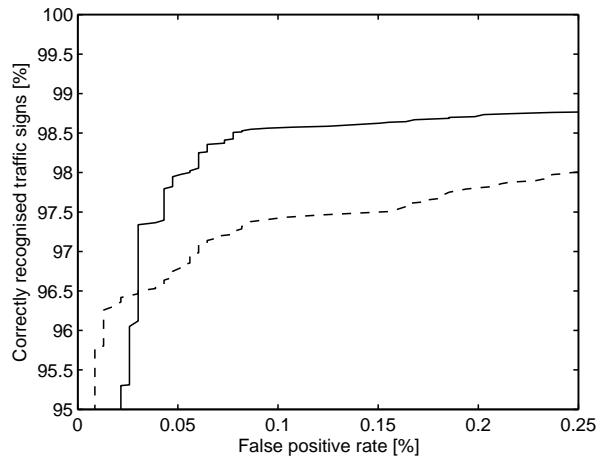


Fig. 3. ROC curve of the classifiers trained with real (solid curve) and synthetically generated (dashed curve) samples.

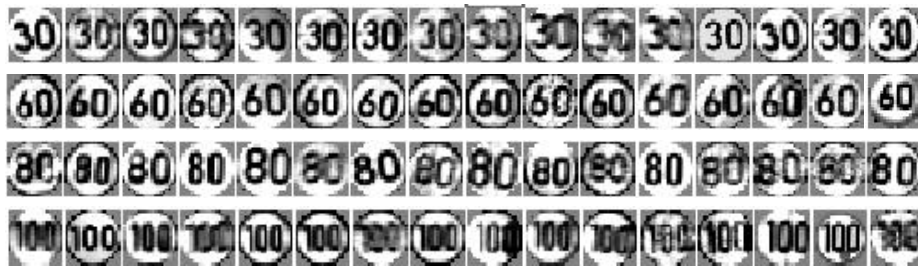


Fig. 4. Selection of traffic sign samples correctly but “nearly” wrongly assigned by the classifier trained with synthetically generated samples. Classifier output is highest for the correct class but only slightly lower for an incorrect class.

rates of less than 0.25%, the rate of correctly recognised traffic signs is only moderately lower for the classifier trained with synthetically generated samples.

Fig. 4 shows several traffic sign samples which are correctly but “nearly” wrongly classified by the classifier trained with synthetically generated samples. Here, the classifier output is highest for the correct class but only a few percent lower for an incorrect class, where we set $c = 1$ for the class assignment defined by Eq. (6). These samples correspond to remarkably strongly tilted, warped, shifted, or blurred traffic signs, thus demonstrating the robustness of the classifier obtained by the synthetic sample generation approach.

To illustrate the difference between the classifier trained with real samples and the one trained with synthetically generated samples, typical samples taken from the test set which are correctly classified by the classifier trained with real



Fig. 5. Typical samples correctly classified by the classifier trained with real samples and incorrectly classified by the classifier trained with synthetically generated samples.

samples and at the same time incorrectly classified by the classifier trained with synthetically generated samples are shown in Fig. 5. For the class assignment, we set $c = 1$ in Eq. (6). For our test set the overall number of such samples amounts to 557. The reason for the misclassification of these traffic sign samples is that the model does not cover the conditions encountered during image formation. For several samples, the rotation angle α is outside the range taken into account by the fit of model parameters. Other samples display a strongly non-uniform illumination across the region of interest, which does not correspond to the utilised photometric model. In some cases a strong blur, e. g. caused by raindrops on the windscreen through which the images are acquired, is apparent. At night or during twilight the samples may display a significant amount of motion blur, caused by the long exposure time, or pixel noise, which is due to the high camera gain. To take into account these samples, it would be necessary to correspondingly extend the transformation model. Another reason for misclassification is the fact that some signs cannot be inferred from the utilised ideal class prototype since e. g. the inlay is horizontally or vertically shifted. To account for these samples it would be necessary to introduce additional, appropriately chosen class prototypes. Furthermore, it may be favourable to apply a bootstrapping stage in order to refine the inferred probability distribution of the model parameters, involving the acquisition and evaluation of misclassified samples analogous to the traditional bootstrapping of classifiers as described e. g. in [8] or [10].

4 Summary and conclusion

In this contribution we have described a synthetic sample generation approach for classification tasks in which each class can be characterised by an ideal pattern, and the variability of the training and test samples is essentially caused by a geometric and photometric transformation of the ideal pattern due to the image formation process. This transformation was assumed to depend on system-specific influencing quantities such as illumination, characteristics of the sensor and optical system, and viewing direction.

We have examined this training approach in detail in the context of a real-world vehicle-based vision system for the recognition of traffic signs. Our image

formation model consists of an affine transform to account for variations in viewing direction and a linear photometric transform defined by a gain and an offset parameter. The model parameters were learned from two different traffic sign classes for which a large number of real samples are available. Based on a Kolmogorov-Smirnov test we have shown at a high confidence level that the empirical parameter distributions inferred from the two traffic sign classes were generated by the same underlying statistical law.

Our experimental system evaluation on a large set of real-world test data demonstrates that the classification rates obtained for classifiers trained with synthetic samples are comparable to those obtained based on real training data, thus clearly demonstrating the usefulness of the proposed approach. Typically, misclassifications of the classifier trained with synthetically generated samples are caused by violations of the model assumptions (e. g. non-uniform illumination, motion blur, pixel noise), by atypical model parameter values, or due to the fact that the samples cannot be derived from the ideal class prototype.

To further increase the classification performance, future work will include a bootstrapping stage to refine the parameter distribution and an extension of the image formation model with respect to the above-mentioned additional influencing quantities.

References

1. H. S. Baird, T. Kanungo, R. M. Haralick. Validation and Estimation of Document Degradation Models. 4th UNLV Symp. on Document Analysis and Information Retrieval, Las Vegas, 1995.
2. H. S. Baird. State of the Art of Document Image Degradation Modeling. DOD-sponsored Symposium on Document Image Understanding Technology, Annapolis, 1999.
3. D. M. Gavrilu, J. Giebel. Virtual Sample Generation for Template-based Shape Matching. IEEE Conference on Computer Vision and Pattern Recognition, vol. I, pp. 676-681, Kauai, 2001.
4. L. Gottesfeld Brown, A Survey of Image Registration Techniques, ACM Computing Surveys 24(4), pp. 325-376, 1992.
5. U. Kreßel, F. Lindner, C. Wöhler, A. Linz. Hypothesis verification based on classification at unequal error rates. Int. Conf. on Artificial Neural Networks, pp. 874-879, Edinburgh, 1999.
6. P. R. Krishnaiah, P. K. Sen (eds.). Handbook of Statistics, vol. 4. Nonparametric methods. North Holland, Amsterdam, 1984.
7. A. Kuhl, L. Krüger, C. Wöhler, U. Kreßel. Training of Classifiers Using Virtual Samples Only. Int. Conf. on Pattern Recognition, vol. III, pp. 418-421, Cambridge, UK, 2004.
8. M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, T. Poggio. Pedestrian detection using wavelet templates. IEEE Conf. Computer Vision and Pattern Recognition, pp. 193-199, San Juan, 1997.
9. J. Schürmann. Pattern Classification. Wiley-Interscience, New York, 1996.
10. C. Wöhler, J. K. Anlauf, T. Pörtner, U. Franke. A Time Delay Neural Network Algorithm for Real-Time Pedestrian Recognition. IEEE Int. Conf. on Intelligent Vehicles, pp. 247-252, Stuttgart, 1998.