A Hierarchical Approach to Sign Recognition

Piyanuch Silapachote and Allen Hanson
Department of Computer Science
University of Massachusetts Amherst
Amherst, MA 01003
{pla, hanson}@cs.umass.edu

Richard Weiss
School of Cognitive Science
Hampshire College
Amherst, MA 01002
rswCS@hampshire.edu

Abstract

Sighted individuals draw a significant amount of information from signs but this information is denied to the visually impaired. VIDI is an evolving system for detecting and recognizing signs in the environment and voice synthesizing their textual contents. The wide variety of signs commonly encountered and the uncontrolled nature of the real world add significant complexity to the problem. VIDI treats the recognition problem as one of matching an unknown sign image, obtained from the detection component as a hypothesized sign, to a database of known signs. A color based support vector machine classifier coarsely picks a group of sign classes that are the most likely matches to the query. A finer retrieval technique employing corners and shape contexts ranks the hypothesized sign classes and verifies whether or not the top ranked class is the true class of the query. The database includes a set of real images with a wide variety of sign classes, each containing multiple signs exhibiting not only illumination differences, but also rotational variations. Tested on over 1,200 images, our system correctly recognizes and identifies the sign class of a query, achieving a 94.75% accuracy.

1. Introduction

An automated sign detection and recognition system provides a visually impaired individual a chance to obtain useful information from signs in much the same way a sighted person does. With an embedded language translator [27], it further eliminates language barriers. Integrated into a Personal Digital Assistant platform, the system becomes a mobile traveling aid [28].

Techniques for sign detection and recognition have recently gained attention from several researchers. However, the limited domain of standardized traffic signs is the main focus in much of the previous work. Sekanina [22], restricted to Norwegian speed limit signs, used a color-based filtering and template matching scheme to locate and read numbers on signs. Liu [14] segmented images via color thresholding and recognized Stop signs using a neural network. Similarly, but in a broader road sign domain, Escalera [9] detected signs using color thresholding and shape analysis, and classified signs by a trained neural network. Another work on color thresholding is that of Bénallal [3] where road signs are detected in real-time.

For a component of the RS^2 colorless sign detection system where edge orientation and hierarchical templates were used to search for specific geometrical shapes of signs [13], Paclík [19] used grayscale information and a moment-based shape descriptor to categorize signs through multiple classifiers in a decision tree. Other approaches include a Laplace kernel classifier by Paclík [20] and a single positioning of a space-variant sensor window by Shaposhnikov [23].

Chen and Yuille [8] developed a visual aid system for the blind that used Adaboost to learn to identify text regions by selecting features with low entropy; an OCR system is used to read or reject the detected text.

Although both text and road signs play an important role in navigation and contain crucial information for drivers and pedestrians, other commercial signs, for example restaurants and banks, are no less significant in people's daily lives. In our work, "sign" is generally defined as any physical sign, including traffic, government, public, and commercial signs. These signs may or may not contain text. Although conceptually approaches to recognizing these signs appear to be similar to those used for traffic signs, recognition of signs under this broader definition is much more difficult and much more challenging. Complications arise not only from how signs are embedded in the environment: their sizes, orientations, and/or occlusions, but also from the extremely broad variations in both text and symbol structures as well as an extremely large set of colors and shapes.

The VIDI (for Visual Integration and Dissemination of Information) for the visually impaired is composed of three stages: sign detection, sign recognition, and speech synthesis. Our solution for the detection task, achieving 74.73%

Proceedings of the Seventh IEEE Workshop on Applications of Computer Vision (WACV/MOTION'05)
0-7695-2271-8/05 $ 20.00 IEEE

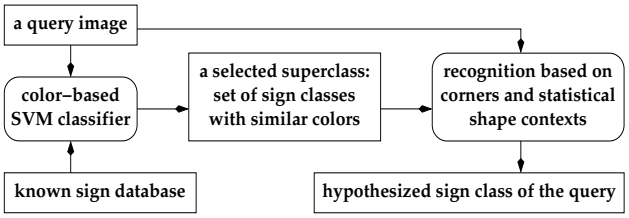

Figure 1. The 2-level hierarchical structure of the proposed sign recognition system.

detection rate and 1.66% false positive rate, utilizes a conditional random field using color and texture features [26]. The features are based on multiscale, oriented band-pass filters, and nonlinear grating cell model filters. These features have been shown to be effective at detecting signs in unconstrained outdoor images. In the recognition system, which is the main focus of this paper, hypothesized sign patches from the detection are matched against a sign database in two stages. First, color information is used to restrict the number of sign classes to be considered. Second, feature points are used to match the query patch with known sign classes in the database. This results in a ranking of the database signs with respect to their similarity to the query and hence a ranking of possible sign classes for the query. From this class the sign content is determined and conveyed to the user via speech synthesis.

In this paper, we address the sign recognition problem in a general domain. Our proposed 2-level hierarchical framework is outlined in Figure 1. A Support Vector Machine (SVM) learns to coarsely classify signs based on their apparent colors. This coarse classification is followed by a finer search using interest points (here corners) and local statistical shape information around the points.

Colors, corners, and shape contexts reinforce one another. While colors may not be present or usable all the time and may be highly affected by natural lighting variations when they are present, corners and shape contexts provide a more stable set of features. Furthermore, corners and shape contexts capture spatial information, which is lost in the construction of color histograms.

Although it might appear that a search via a correspondence of corners and associated contexts might be sufficient to solve the problem, the high computational complexity of this technique makes it difficult to apply corner matches to a large database. For this reason, the color-based SVM classifier, which quickly narrows down a search, is an important focus of attention mechanism.

2. Color-Based SVM Classifier

Humans are remarkably good at adapting to most color variations that occur in the real world, including (but not limited to) lighting effects, reflections, refractions, specularity, and shadowing. In computer vision, color as a feature works well in restricted domain but becomes very unstable in natural uncontrollable settings. In order to overcome these phenomena and to build more stable color models, Buluswar [5] constructed a model of the measured color of daylight for a broad range of sky conditions and adapted the Dichromatic Reflectance model to use the daylight model. This approach relies heavily on detailed weather and sky conditions when the sample measurements are taken. Miller [15] developed a flow-based statistical model of color changes that work quite well in predefined indoor settings.

Under outdoor imaging with unspecified conditions, in contrast, it is virtually impossible to eliminate illumination effects or to firmly establish color constancy. On the other hand, throwing away very distinctive color information is not desirable. As a compromise, we use color information for a coarse classification of signs to quickly narrow down the search space to a few possibilities, which are then further explored using more powerful features.

We represent an image using a color feature vector [24], which is a concatenation of two 1-dimensional histograms corresponding to hue and saturation in the hue-saturation-value (HSV) color model. The value component is ignored because it relates to scene brightness which varies considerably. While hue characterizes the color of a pixel, saturation reflects how much pure hue mixes with white light. The contribution of the pixel colors, i.e. hue values, to the hue histogram should vary proportionally to their purity, stability, and reliability, i.e. saturation values. This observation is naturally encoded during the histogram construction by weighting the contribution of a pixel to the hue histogram by its saturation value (as opposed to a pixel count as commonly used). Furthermore, only pixels with a saturation value above a predefined threshold contribute to the hue histogram. This is to account for the undefined nature of hue when there is insufficient color information (grayscale).

Once color feature vectors are formed, a SVM classifier learns to coarsely classify signs, narrowing down the search to a few classes with similar colors. SVM is chosen as a classifier due to its ability to generalize well. SVM is an approximate implementation of the Structural Risk Minimization induction principle [18] in which a generalization error is minimally bounded and a margin is maximized. Formulated for dichotomy, a multi-class categorization is handled by a one-against-all voting scheme [10].

3. Corners and Shape Contexts

Another useful discriminating feature is shape. A key question is what shapes are and how to identify them. In a traffic sign domain only a few shapes are typically found, which can be easily modeled using standard techniques, such as the Hough transform [25]. In the general case, signs can be of arbitrary shape and so it is neither possible nor practical to model them with a set of discrete well defined shapes. We explore the use of a statistical shape descriptor that represents an object by a set of points; from these points local shape information is inferred from the characteristic of this point set. Shape information is determined by the geometric relations between the points as a whole [12].

Shape contexts have been proposed and successfully applied to a number of applications [2, 17] exhibiting both rigid and non-rigid transformations. However, in most cases, the images are fairly clean and salient points can be accurately selected so the registration between a pair of images can be well established. Unfortunately, this is not usually the case in outdoor images, where occlusions and lighting conditions make the point correspondence/registration problem difficult.

We extend the shape context design so that it is more suitable to the complex outdoor sign recognition task. Corners are used as salient points; these are extracted using a detector that is based on an analysis of local anisotropism [6]. The shape context is modified so that it is both rotation and scale invariant. The final matching cost between any image pair is computed through a correspondence problem, an attempt to map a corner point set of one image onto the other based on characteristics of their shape contexts.

3.1. Corner Detection

Since a corner marks an intersection of lines, every corner point lies on edges. Instead of an original image, we detect corners on a grayscale image whose brightness represents edge strength computed by a Canny edge detector [25]. Another advantage of this is a reduction of noise that transfers to an increased accuracy in corner detection.

In the detection of corners, we modify the definition given in [6] where corners are points with strong gradient intensity and without a single dominant gradient orientation. The cornerness $c(\vec{x})$ of any point $\vec{x} \in \Re^2$ on an intensity image I is defined as

$$c(\vec{x}) = \begin{cases} \|\nabla I(\vec{x})\| & \text{for} \quad g(\vec{x}) < 0.5 \\ 0 & \text{for} \quad g(\vec{x}) \geq 0.5 \end{cases} \tag{1}$$

$$g(\vec{x}) = \frac{(\iint_\Omega (I_x^2 - I_y^2) dxdy)^2 + (\iint_\Omega 2 I_x I_y dxdy)^2}{(\iint_\Omega (I_x^2 + I_y^2) dxdy)^2} \tag{2}$$

$$\theta(\vec{x}) = \tan^{-1} \frac{\iint_\Omega 2 I_x I_y dxdy}{\iint_\Omega (I_x^2 - I_y^2) dxdy} \tag{3}$$

$g(\vec{x})$ and $\theta(\vec{x})$ are the strength and the orientation of the anisotropism of a pattern within a small neighborhood Ω centered at a point $\vec{x}$. $\|\nabla I(\vec{x})\|$ is a gradient magnitude at $\vec{x}$. The range of an inverse tangent function is defined for $[-\pi, \pi]$ radians.

Post-processing on a cornerness map includes non-maximum suppression [11, 25]. Ridges around local maxima are thinned by identifying along the direction of the gradient a pixel with the maximum gradient magnitude. This finds a corner center producing a 1-pixel wide corner feature point.

Additionally, every corner has at least two arms (associated edges) and should have sufficient support from its arms. To incorporate this, we measure and threshold $\psi_c(\vec{x})$, the likelihood of neighboring pixel $\vec{x}$ of a corner c being arms of c [6]. α is an angle between $\theta(\vec{x})$ and a directional vector from c to $\vec{x}$.

$$\psi_c(\vec{x}) = g(\vec{x}) \cdot \|\nabla I(\vec{x})\| \cdot \cos^3 \alpha \tag{4}$$

3.2. Shape Context

Assume n salient points are detected. A shape context associated with any sample point is a distribution of relative positions from the point itself to all other $n-1$ points. Since the descriptor should be more sensitive to neighboring pixels than to those further away, histogram bins are uniformly quantized in a log-polar space.

Mathematically, a shape context at a point p_i is a histogram whose k^{th} bin equals the cardinality of the set

$$\{p_j \neq p_i : [\angle(p_i, p_j) \text{ and } \log(dist(p_i, p_j))] \in bin(k)\} \tag{5}$$

$$dist(p_i, p_j) = \frac{p_j - p_i}{\max_{\forall l,m \in [1,n]} (p_m - p_l)} \tag{6}$$

where $i, j \in [1, n]$, $\angle(p_i, p_j)$ and $p_i - p_j$, respectively, are an angle and a Euclidean distance between points p_i and p_j. The denominator in $dist(p_i, p_j)$ is a normalizing factor that makes the shape context scale invariant.

3.3. Solving the Correspondence Problem

Let a shape context be denoted by a histogram h and let $h(r, s)$ denote a value at a bin (r, s) of h that accounts for neighboring points within rotation angle $r \in R$ and distance $s \in S$ where R bins are used along an angular dimension and S bins along a radial distance.

$$h^t(r, s) = h(mod(r + t, R), s) \tag{7}$$

defines a shift t of a histogram h, where $t, r \in [0, R-1]$ and $s \in [0, S-1]$. By construction, a t-shifted shape context is equivalent to a shape context of an image physically rotated by $t \cdot (2\pi/R)$ degrees counterclockwise. The distance $C_{i,j}^t \equiv C^t(p_i^t, q_j)$ with respect to a shift t between

two shape contexts, p_i at a point on the first image and q_j at a point on the second image, is measured based on normalized cross correlation.

$$M = \min_t\{\min_\sigma[\Sigma_{i=1}^n C_{i,\sigma(i)}^t]\} \quad (8)$$

$$t^\star = \arg\min_t\{\min_\sigma[\Sigma_{i=1}^n C_{i,\sigma(i)}^t]\} \quad (9)$$

$$\sigma^\star = \arg\min_\sigma[\Sigma_{i=1}^n C_{i,\sigma(i)}^t] \quad (10)$$

M is the similarity measure between an image pair, n is the total number of salient points, and σ a permutation of points on the second image corresponding to a one-to-one mapping to points on the first image. $\sigma^\star$ is the permutation that leads to minimum matching cost. $t^\star$ represents an optimal shift, which implicitly states that bringing the two images in a pair to the best alignment requires an approximate $t^\star(2\pi/R)$ degree rotation of one image.

The minimization of the matching cost over all possible permutations is an instance of the square assignment problem. One approach to solving this is the Hungarian technique [21] with a computational complexity of $O(n^3)$. An implementation by Borlin [4] is used here.

The approach described above is applicable when both images in a pair have an equal number of salient points. However, our detector returns all satisfying corners, which may result in a different number of salient points for each image. Our solution is to utilize dummy corners in order to even out the number of points. As a consequence, extra checks are required to add and drop these dummies for the final matching result.

4. Experiments and Results

4.1. Data Collection

Images of signs were taken around downtown Amherst, MA, using a still digital camera (Nikon Coolpix 995) with automatic white balance on. Spot metering was used along with manual +/- exposure adjustment to control the amount of light projected onto the camera sensor.

We capture natural lighting and illumination effects by taking 5 original images of each perceptual sign class from the same physical location at five different times of the day, that is, approximately 2-3 hours apart beginning at daybreak and finishing at nightfall. The effect of the image background was eliminated by manually segmenting the images. Every original image is a frontal view. To obtain images at multiple in-plane rotations, each original image was rotated from -90° to 90° with a 10° interval (except at 0°, which is the original image). Figure 2 shows an example of one sign image; its original image along with its segmented and rotated images.

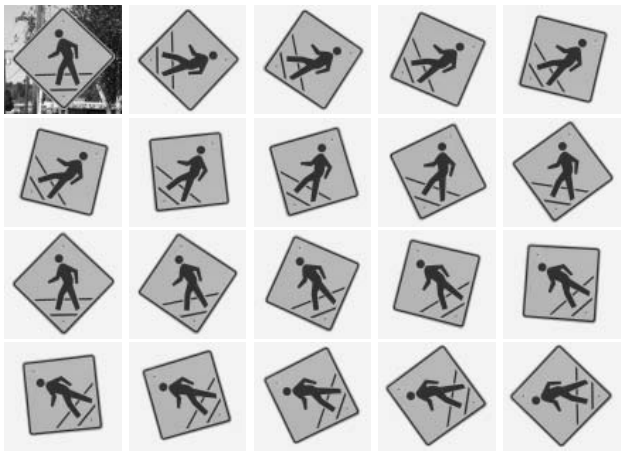

Figure 2. A data sample showing one original image, its segmented, and rotated images.

Figure 3. Sample of prototypes of the six sign classes in the blue-colored superclass.

A database consisting of 24 sign classes for a total of 2280 sign images has been constructed and is available on the VIDI web page [1]. The 24 classes are divided by similarity of colors into 8 superclasses. A brown-colored and 3 other multi-colored signs make up four single superclasses. The other four superclasses contain 2, 5, 6, and 7 individual sign classes, corresponding to yellow, red, blue, and green colored signs. A sample prototype of every sign class in the blue superclass is shown in Figure 3.

4.2. Color-Based SVM Classifier

Color histograms are constructed for every image (see Figure 4 for samples of extracted color histograms) and used as color feature vectors for an SVM classifier [7] with a Gaussian Radial Basis kernel. A saturation threshold was arbitrarily set at 0.1 where saturation values range from 0 to 1 inclusively. 64 bins are used for hue and 10 for saturation. These bins are coarsely defined partially to account for possible shifts of colors due to illumination effects and other noises. Since the classifier relies only on colors, clearly it is unable to distinguish signs whose colors are similar. Two signs are similar under the HSV color feature when both

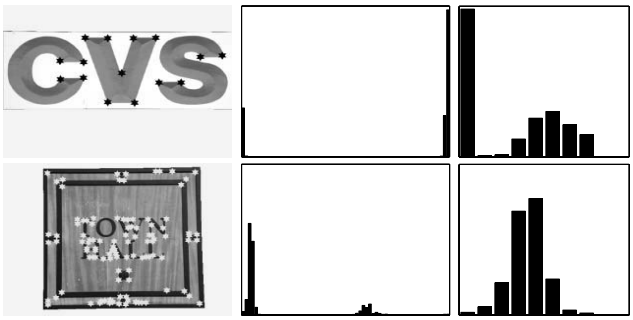

Figure 4. Image features for the CVS sign are shown on the top row and the Town Hall sign on the bottom row. The first column shows detected corners overlay on the sign images. The second and third columns, respectively, are the hue and the saturation histograms where the x-axis represents the histogram bins from red to magenta for the hue and from low to high values for the saturation.

contain equal proportions of the same colors. Accordingly, we divide all signs in the database into superclasses, each of which consists of classes exhibiting similar colors.

The goal of the color-based SVM is to learn to classify signs according to their defined superclasses. Our statistical results on the performance of the SVM is based on a 5-fold cross validation analysis [16], where the classifiers are trained on 80% of the data and tested on the other 20% in 5 runs, each run retaining different subsets (20% in each case) of the data as the test set. The classification accuracy on the test data portion is 100%.

4.3. Matching of Corners and Shape Contexts

Among all superclasses, those which contain only a single sign class are completely classified by the SVM. Other query images that fall into the other superclasses, which contain multiple classes, are passed from the SVM to the corner and shape context matching process comprising the second stage of the recognition system.

Given a query image and a hypothesized superclass, the matcher attempts to single out the one class that is the most probable match to the query. Let a prototype be a representative image of a class in a conjectured superclass; specifically, the 5 non-rotated original images of each class in the database are our prototypes. On every prototype and a query image, corners are detected (see Figure 4 for samples of sign images with their detected corners marked) and shape contexts are extracted (8 bins are defined along the angular dimension and 16 along the radial distance). A pair-wise comparison between a query and every prototype is performed. All prototypes are then ranked in increasing order of their match costs against a query. Implicitly this is a ranking of corresponding sign classes in order of decreasing likelihood of being the true class of the query. The most probable class, ranked first, is hypothesized as the one to which a query image belongs.

Testing on 1,200 queries randomly chosen from a pool of sign images of the 20 classes that need to be passed from the first to the second stage of the system, the matcher achieves 94.75% accuracy looking at the first rank alone. A sample of the ranking is shown in Figure 5. The query is the image in the upper left corner. The remaining images are those returned by the matcher in order of the match cost (shown by the bar below each image); recall that low match costs are good. In this case, the query is highly rotated at 90° angle. The five prototypes in the correct class corresponding to the query are ranked 1 through 5. The second candidate is another street-name sign whose context is very closed to that of the query, but the matcher successfully distinguishes them with a significant difference in the match costs.

Figure 6 shows the average match scores by rank over all test queries for which the true class of the query appears at the first rank. Note that the average scores exhibit a sharp change between the 5^{th} and the 6^{th} rank. This change is not apparent in Figure 7, which is the same as Figure 6 except that in this case the matcher has failed to find the correct class of the query in the first rank (that is, an incorrect class appears in the first rank). What this tells us is that the correct class is well separated from the others by the matcher (recall that there are 5 prototypes of each class in the comparison). That is, the query matches well only with its true class and that most of the time all of the 5 prototypes come up at the first 5 ranks. On the other hand, in misclassified cases, the queries have a hard time matching with both their true class and the other prototypes resulting in a much higher match costs across the board.

Figure 8 plots the cumulative classification accuracy of the matcher where a query is said to be correctly classified at rank k if at least one of the prototypes of its true class turns up within the first k ranks in the list. As shown in the graph, the matcher reaches 100% accuracy at rank 10. What this means is even though the matcher fails to classify some of the queries in the first rank, the correct class is represented in the first 10 ranks. Instead of restricting the classification decision to only the first rank, one could consider the first few ranks (1 to 10 in our case) and perform a more refined matching process to verify the real class of the query within these few prototypes. These results illustrate the performance of the matching process while the emphasis is on a search for the best possible sign class. Our next step will include the rejection of queries when their corresponding classes are not represented in the database.

5. Conclusions and Future Work

A 2-level hierarchical system for sign classification and recognition was proposed and tested successfully on an experimental database. The first stage of the recognition system uses global color features and learns via a support vector machine to coarsely classify signs. The second stage utilizes global spatial relationships among local corners, encoded as shape contexts, to refine the search and hypothesized a single class of sign as the end product. The system has achieved, based on the experimental dataset, a near perfect classification accuracy and a robustness in tolerance to image rotations and illumination variances.

VIDI is a work in progress. One immediate goal is to extend the dataset to a larger collection of sign classes. This extension not only includes a larger set of sign classes but also a more complicated variation of sign images within a single class. An influence of occlusion is to be explored; an effect of a perspective distortion is to be considered, as well as that of a controllable amount of motion blur. The increased complexity of the expanded dataset may well reduce the SVM accuracy, resulting in erroneous classes being forwarded to the matcher. At that point, additional steps will be needed to handle the propagation of errors from one level of the hierarchy to the next. A second goal is to reduce the computational complexity of the matching process.

Figure 5. An example of ranking results of the matcher. The image in the upper left corner is the query, the remaining 19 images are the matches in rank order. The length of the bar on the bottom of each sign represents the match score.

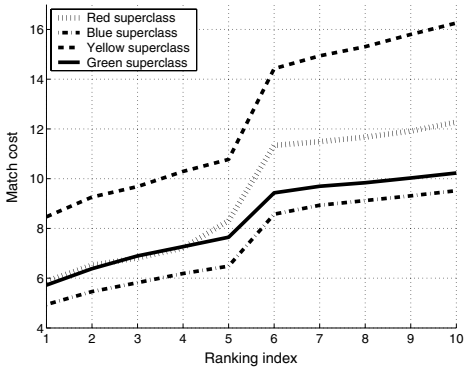

Figure 6. The average match scores by rank of every query correctly classified.

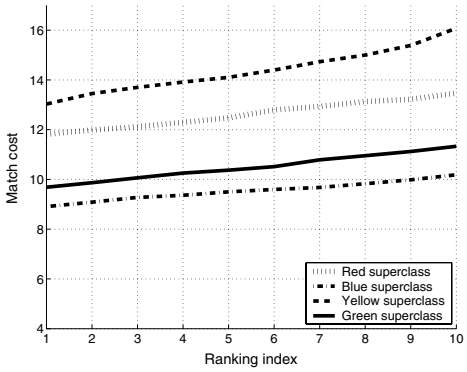

Figure 7. The average match scores by rank of every misclassified query.

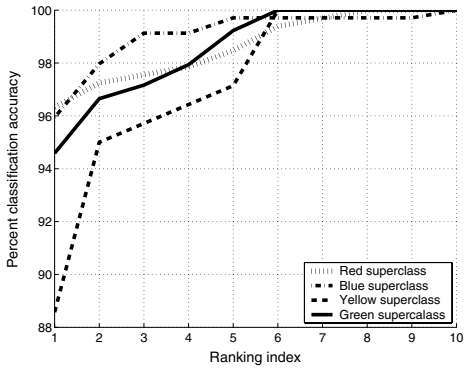

Figure 8. Cumulative classification accuracy of the matcher.

Acknowledgments

The authors would particularly like to thank Marwan Mattar for his collection of the sign images. In addition, we would like to thank Andrew Barto for his useful discussions and valuable suggestions. This work is supported by the NFS grant IIS-0100851.

References

[1] Vidi project web page. Univ. of Massachusetts Amherst, http://vis-www.cs.umass.edu/projects/vidi/index.html.

[2] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 24(4), 2002.

[3] M. Bénallal and J. Meunier. Real-time color segmentation of road signs. In *Proc. IEEE Canadian Conf. on Electrical and Computer Engineering*, Montréal, Québec, CA, May 2003.

[4] N. Borlin. Matlab implementation of the Hungarian. Dept. of Computing Science, Umeå University, Sweden, 1996.

[5] S. D. Buluswar. *Apparent Color in Outdoor Images*. PhD thesis, University of Massachusetts, Amherst, MA, 2002.

[6] F. Chabat, G. Yang, and D. Hansell. A corner orientation detector. *Image and Vision Computing*, 17(10), Aug. 1999.

[7] C. Chang and C. Lin. Libsvm - A Library for SVM, 2002.

[8] X. Chen and A. L. Yuille. Detecting and reading text in natural scenes. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Washington, DC, June 2004.

[9] A. de la Escalera, L. Moreno, M. Salichs, and J. Armingol. Road traffic sign detection and classification. *IEEE Transactions on Industrial Electronics*, 44(6):848–859, Dec. 1997.

[10] T. Dietterich and G. Bakiri. Solving multiclass learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research*, 2:263–286, Jan. 1995.

[11] D. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Prentice Hall, New Jersey, 2003.

[12] C. Grigorescu and N. Petkov. Distance sets for shape filters and shape recognition. *IEEE Transactions on Image Processing*, 12(10):1274–1286, Oct. 2003.

[13] V. Líbal, T. Zikmund, P. Paclík, B. Kovář, Králík Martin, P. Zahradník, and M. Vlček. Traffic sign identification and automatic template generation. In *CTU Workshop*, 1996.

[14] H. Liu and B. Ran. Vision-based Stop sign detection and recognition system for intelligent vehicles. Transportation Research Record No.1748:161-166, 2001.

[15] E. G. Miller and K. Tieu. Color eigenflows: Statistical modeling of joint color changes. In *Proc. of the IEEE Intl. Conf. on Computer Vision*, Vancouver, B.C., Canada, July 2001.

[16] T. Mitchell. *Machine Learning*. WCB/McGraw-Hill, 1997.

[17] G. Mori and J. Malik. Recognizing objects in adversarial clutter: Breaking a visual captcha. In *Proc. IEEE Computer Vision and Pattern Recognition*, Madison, WI, June 2003.

[18] E. Osuna, R. Freund, and F. Girosi. Support vector machines: Training and applications. A.I. Memo 1602, Artificial Intelligence Laboratory, MIT, 1997.

[19] P. Paclík and J. Novovicová. Road sign classification without color information. In *Proc. of the 6th Annual Conference of the Advanced School for Computing and Imaging*, Lommel, Belgium, June 2000.

[20] P. Paclík, J. Novovicová, P. Pudil, and P. Somol. Road sign classification using the laplace kernel classifier. In *the 11th Scandinavian Conf. on Image Analysis*, Greenland, 1999.

[21] C. Papadimitriou and K. Steiglitz. *Combinatorial Optimization: Algorithms and Complexity*. Prentice Hall, NJ, 1982.

[22] L. Sekanina and J. Torresen. Detection of Norwegian speed limit signs. In *European Simulation Multiconference*, 2002.

[23] D. Shaposhnikov, L. Podladchikova, A. Golovan, N. Shevtsova, K. Hong, and X. Gao. Road sign recognition by single positioning of space-variant sensor window. In *Proc. of the 15th International Conference on Vision Interface*, Calgary, Canada, May 2002.

[24] M. Swain and D. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, Nov. 1991.

[25] E. Trucco and A. Verri. *Introductory Techniques for 3-D Computer Vision*. Prentice-Hall, Inc., New Jersey, 1998.

[26] J. Weinman, A. Hanson, and A. McCallum. Sign detection in natural images with conditional random fields. In *Proc. of IEEE International Workshop on Machine Learning for Signal Processing*, pages 549–558, São Luís, Brazil, Sep. 2004.

[27] J. Yang, J. Gao, Y. Zhang, X. Chen, and A. Waibel. An automatic sign recognition and translation system. In *Proc. Workshop on Perceptive User Interfaces*, Orlando, FL, Nov. 2001.

[28] J. Zhang, X. Chen, J. Yang, and A. Waibel. A PDA-based sign translator. In *Proc. of the International Conference on Multimodal Interfaces*, 2002.