

Hybrid ensemble of classifiers for logo and trademark symbols recognition

Bogusław Cyganek

© Springer-Verlag Berlin Heidelberg 2014

Abstract The paper presents a hybrid ensemble of diverse classifiers for logo and trademark symbols recognition. The proposed ensemble is composed of four types of different member classifiers. The first one compares color distribution of the logo patterns and is responsible for sifting out images of different color distribution. The second of the classifiers is based on the structural tensor recognition of local phase histograms. A proposed modification in this module consists of tensor computation in the space of the morphological scale-space. Thanks to this, more discriminative histograms describing global shapes are obtained. Next in the chain, is a novel member classifier that joins the Hausdorff distance with the correspondence measure of the log-polar patches computed around the corner points. This sparse classifier allows reliable comparison of even highly deformed patterns. The last member classifier relies on the statistical affine moment invariants which describe global shapes. However, a real advantage is obtained by joining the aforementioned base classifiers into a hybrid ensemble of classifiers, as proposed in this paper. Thanks to this a more accurate response and generalizing properties are obtained at reasonable computational requirements. Experimental results show good recognition accuracy even for the highly deformed logo patterns, as well as fair generalization properties which support human search and logo assessment tasks.

Keywords Logo recognition · Trademark classification · Structural tensor · Modified Hausdorff distance · Fuzzy image measures

Communicated by V. Loia.

B. Cyganek (✉)
AGH University of Science and Technology, Al. Mickiewicza 30,
30-059 Kraków, Poland
e-mail: cyganek@agh.edu.pl

1 Introduction

Logo and trademark recognition concerns classification of usually artificially created images which content typically reflects name of a company, an organization, goods for sell, etc. The form of a logo or a trademark is commonly designed to attract people, and therefore they constitute a specific group of images. Frequently, these are characterized by specific colors, contrast or other attributes. Although such features seem to be distinctive, at least in the light of, for example, outdoor images, their recognition is still far from being perfect. Accuracy of the known techniques depends on the quality of the processed images, their resolution, as well as geometric transformations and noise which affected the logos contained within. Even more problematic is logo and trademark recognition in the search tools which response time has to be limited to fraction of seconds.

In this paper we address this problem, presenting logo and trademark recognition system which can be characterized by high accuracy, good generalization properties, and real-time performance. Apart, the system architecture is based on plug-in modules which can be flexibly configured in either order. Such interface allows setting serial, parallel or even hybrid output-input like connections of the classifiers. In other words, the classifiers can operate in parallel, providing a common recognition rate, or they can be set into a chain in which output of the predecessor is put as an input to its successor. Also, some of the plug-ins can be turned off or new types of plug-ins can be added if necessary. Such flexible architecture allows high recognition accuracy which compares or in some cases outperforms the known solutions, as will be discussed. Also, thanks to the used fast image processing methods, the whole system can perform in real-time, which is one of requirements of modern search tools.

High accuracy of the system is achieved thanks to diversity of member classifiers composing the system. In this role we propose four different classifiers. The first one is based on computation of the structural tensor (ST) which provides information on low-level content of the images. Then, the global shape is recognized comparing histograms of the orientation phases of the local patterns in the images. However, the method is further extended by computation of the ST in the morphological scale-space. This approach allows extraction of more discriminative information and well reflecting content of the processed images, as will be discussed. The second classifier is based on computation of the Hausdorff measure which compares geometrical configuration of point sets, characteristic to the processed patterns. In our system, these point sets for Hausdorff measure are the corner points, which are easily found analyzing eigenvalues of the already computed ST. Additionally, in this paper we propose a novel modification of this matching scheme which consists of adding to the Hausdorff distance a correspondence measure between the image patches processed in the log-polar (LP) domain. The main idea behind this method is comparison in the LP domain, in which changes of rotation and scale transform into easily detectable horizontal and vertical shifts of the patterns. On the other hand, a third of the proposed member classifiers utilizes the statistical moments to compute the affine moment invariants (AMI), which allow recognition of the affinely transformed objects. However, this module works poorly for occluded patterns or patterns greatly affected by noise, as will be discussed. The last, fourth classifier, is used to sift out images which comply with a specific color distribution. This is done by comparison of histograms of color channels, although other advanced methods can be also used in this respect (Cyganek 2013; Cyganek and Wiatr 2011).

Each of the aforementioned classifiers has its advantages and disadvantages, and can perform with different accuracy based on the contents and quality of the processed pattern. However, as will be shown and which is our main contribution, the above classifiers when joined together into a hybrid ensemble of cooperating classifiers, can greatly enhance accuracy of the recognition process, showing good generalization properties at the same time. Apart from this, computational and memory requirements are low which makes the recognition process robust. Last but not least, the member classifiers can be freely used also in commercial logo recognition systems. In the experimental task, the system was checked in two directions. The first is to measure robustness of the system, expressed in terms of an overall accuracy and response time, to the various changes of the input patterns. In this respect the proposed system shows very good properties, as will be shown. The second answered question regards measurement of the generalization properties of the system. This is to discover an ability of the system to output similar

patterns which seem also similar to the persons performing the classification task of the logo and trademark patterns. The last property is very desirable for practical applications of the presented system.

The rest of the paper is organized as follows: In Sect. 2 related works on different logo and trademark detection and recognition systems are discussed. Section 3 presents a general view on architecture of the presented ensemble of hybrid member classifiers. In Sect. 4, the base logo recognition methods are described. In this section details of all of the aforementioned member classifiers are presented. Section 5 deals with fusion method of the member classifiers. Section 6 presents experimental results with discussion of the obtained results. Conclusions are provided in Sect. 7. The paper ends with bibliography.

2 Related works

Various logo and trademark detection and recognition systems are reported in literature. These can be further divided based on different types of processed images, as well as conditions of operation. In this respect we have few distinctive groups. The first group of methods belongs to the domain of document processing (Chen et al. 2003), the second is logo and trademark recognition in outdoor conditions (Gori et al. 2003), such as sport events (Ballan et al. 2008), the next one concerns car logo detection and recognition (Dai et al. 2009; Pysillos et al. 2010; Mao et al. 2013), just to name a few.

The complete systems are frequently arranged as detection followed by the recognition stage (Cyganek 2013). In respect to the former, Pham proposes an unconstrained logo detection in document images which is based on segmentation and subsequent calculation of the spatial density of the predefined foreground pixels (Pham 2003). On the other hand, in the paper by Chen et al. (2003) a method for noisy logo recognition using line segments and Hausdorff distance is proposed. Their method is based solely on processing of line segments. Then, the dissimilarity measure is computed between sets of line segments rather than sets of points. Gori et al. (2003) propose a logo recognition system based on a modified backpropagation neural network. Their edge-backpropagation is a modified learning algorithm derived from the classical backpropagation by addition of a weighting error function in place of the Euclidean norm. The weights of this function are computed based on a gradient of an image which accelerates texture rich areas. Color logos and trademark recognition method for content based image retrieval (CBIR) system is proposed in the paper by Phan and Androutsos (2010). Their method is based on application of the color edge co-occurrence histograms for object detection. As reported, the color edge detection operating in the HSV color space, when combined with the vector order

statistics allows a more accurate representation of edges as compared to the standard algorithms. In other work, [Diligenti et al. \(2001\)](#) presented an adaptive graphical pattern recognition method for the classification of company logos. They exploit the idea of highly structured representation of patterns processed by recursive neural networks which perform classification using connectionist-based learning algorithms. As reported, the method is appropriate for company logo recognition corrupted by noise. A logo retrieval system based on the speeded-up robust sparse features (SURF—[Bay et al. 2008](#)) is proposed in the paper by [Jain and Doermann \(2012\)](#). A filtering method is also proposed using the orientation of local features, as well as their geometrical constraints.

A separate group of research activities concerns the problem of logo and trademark detection in real images. For instance [Ballan et al. \(2008\)](#) propose a method for automatic detection and recognition of trademarks in sport videos. In their method, the scale-invariant feature transform (SIFT) based features were employed, which after matching and clustering of the matched feature points are used to localize a trademark in a frame. Answer of the system is further based on the similarity threshold which is learned by the system. Another active domain is vehicle logo recognition, usually aimed at road surveillance. For instance, in the recent work, [Mao et al. \(2013\)](#) propose a rapid vehicle logo region detection method based on information theory. Their method starts with computation of the image gradients, followed by formation of the binary image based on the saliency map obtained from the gradients. Finally, regions of maximal information content are searched. This method can be used as an image preprocessing step for further car logo recognition. For this purpose, the system proposed by [Psylos et al. \(2010\)](#) can be

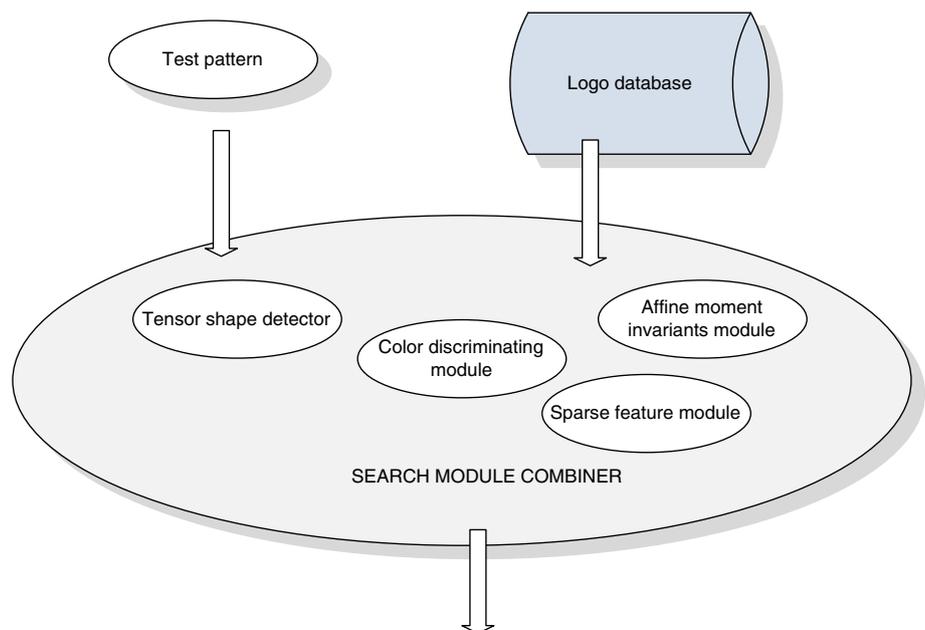
used. The method relies on object recognition with sparse features descriptors. However, computation burden in this case is high and the method requires setting of many parameters. On the other hand, [Petrovic and Cootes \(2004\)](#) demonstrated a vehicle logo recognition system which operates on edge gradients with a refined matching algorithm. Nevertheless, the obtained results are sensitive to viewpoint change as well as plane rotation. In other work, [Dai et al. \(2009\)](#) propose application of the Tchebichef moment invariants and SVM [Dai et al. \(2009\)](#). However, these are also sensitive to noise and geometrical distortions.

Recently, the ensembles of sometimes weak classifiers showed their superiority when compared to single classifier based solutions. An example serves the method by [Viola and Jones \(2004\)](#) for face detection. In their method, the Haar wavelets are used as low-level features in a chain of simple classifiers to select from background pixels which belong to human faces. An overview of ensemble methods is presented in the book by [Kuncheva \(2004\)](#), as well as in the book by [Woźniak \(2013\)](#). Also, an interesting introduction to the ensembles of classifiers is the paper by [Polikar \(2006\)](#). Finally, descriptions of specific realizations and applications of methods from this group can be found in many papers, for instance in the works ([Cyganeck 2013](#); [Cyganeck and Wiatr 2011](#); [Krawczyk et al. 2014](#); [Lughofer and Buchtala 2013](#)), to name a few.

3 System architecture

Figure 1 depicts architecture of the proposed hybrid system for logo and trademark pattern recognition. The processing

Fig. 1 Components of the system for logo and trademark recognition. The processing path includes a number of base classifiers which are set in an ensemble of cooperating modules. They compare the test pattern with the patterns from the logo database. The purpose of this query is to provide the most similar patterns to the test one, as well as the fit measure



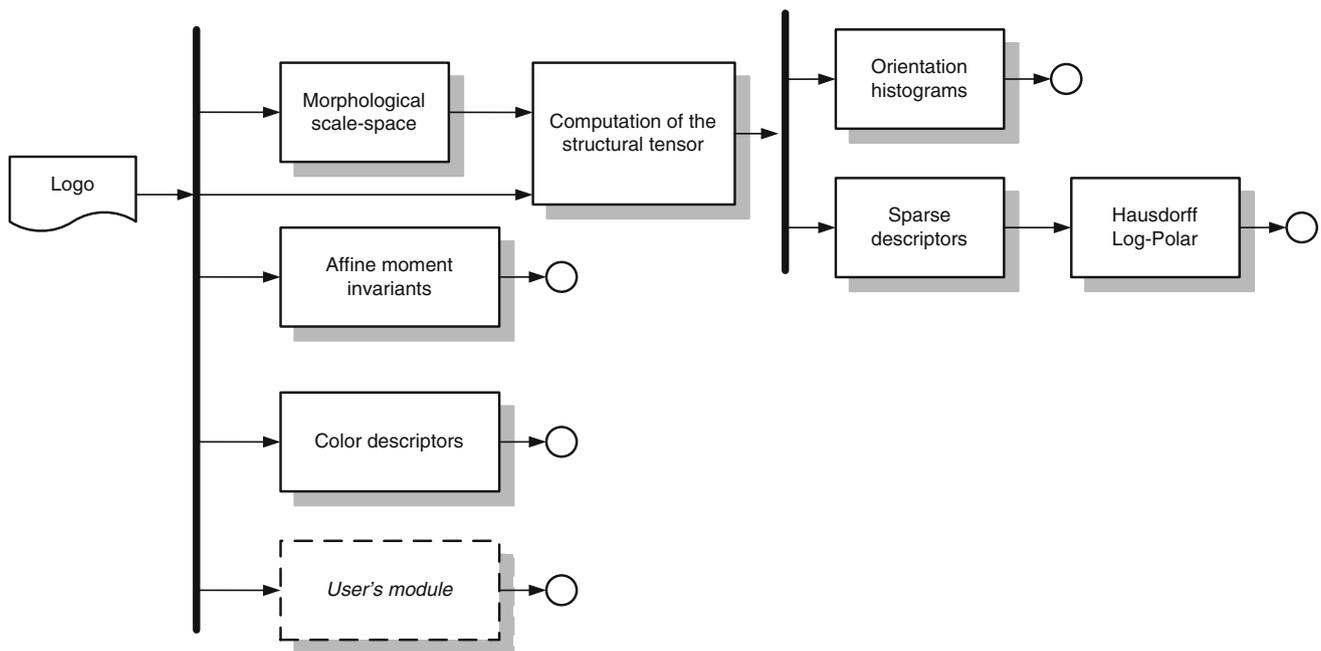


Fig. 2 Activity diagram of the proposed system

path includes a number of base classifiers which are set in an ensemble of cooperating classifiers.

The classifiers answer class of a test pattern based on information gathered from the training patterns, contained in the logo database. The purpose of this query is to provide the most similar patterns to the test one, as well as quantitatively measure of their resemblance ratio, providing a suitable match measure. Figure 2 shows activity diagram of the proposed recognition system. The system is highly flexible and allows joining of the base search modules into different configurations. For instance, all classifiers can be set in series, or they can be arranged into a parallel structure.

Also, any hybrid connection is possible, for example placing the color discriminating module, followed by a set of classifiers organized in a parallel arrangement. It is also interesting to observe that activities of the parallel modules can be performed independently. This feature greatly increases robustness of the proposed system, especially when run on systems endowed with multi-core or on GPU processors (Bugaj and Cyganek 2012).

Also interesting is to observe that the low-level computation of the structural tensor serves two subsequent modules. The first is the orientation histogram computed in the morphological scale-space. However, the second one computes the corner points, which are then used in computation of the Hausdorff distances between logo images.

The proposed architecture is open. That is, other new classification modules can be added. Also, inputs and outputs of the classification modules can be joined in any order, as will be presented in the experimental section of this paper.

4 Base classification methods

4.1 Finding local structures with the structural tensor

The structural tensor (ST) allows low-level analysis of image contents which allows finding the so-called structural places (Bigun 2006). These are image regions which follow some orientation pattern. In contrast, noisy regions don't show any regularity. Both types of situation can be detected with help of the ST. The main idea behind this technique is to represent a local neighbourhood Ω with a single orientation vector \mathbf{w} which fits best to the contents of Ω . This is accurate if most of the gradients in Ω coincide with \mathbf{w} . Computationally, this objective can be achieved by minimizing a cumulative sum of the residual vectors \mathbf{s} in Ω (Cyganek and Siebert 2009).

As already mentioned, the main idea behind the ST is to find a way of representing a local neighbourhood with a representing vector \mathbf{w} , as shown in Fig. 3. This is possible if local regions show some regularity of its intensity signal. Such conditions can be assessed computing intensity gradient at each location. If the gradients are oriented in direction of \mathbf{w} , then \mathbf{w} represents well a given local neighbourhood. Mathematically, this condition can be measured analysing squared modules of vectors \mathbf{s} , shown in Fig. 3, as follows (Brox et al. 2006)

$$\mathbf{s} = \mathbf{g} - \mathbf{r} = \mathbf{g} - \frac{\mathbf{w}}{\|\mathbf{w}\|} \|\mathbf{r}\|. \quad (1)$$

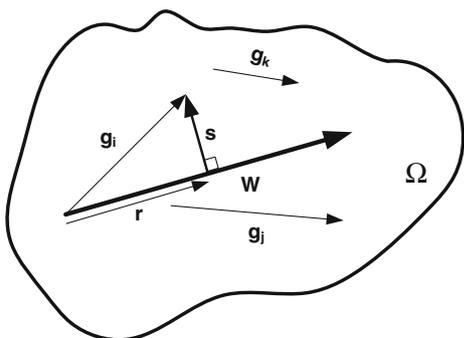


Fig. 3 An illustration of the local neighborhood Ω of pixels in a digital image. The structural tensor aims at representing Ω with a single orientation vector \mathbf{w} . Such representation is accurate if most of the gradients in Ω coincide with \mathbf{w} . A condition for this property can be checked by minimizing a cumulative sum of the residual vectors \mathbf{s} in Ω

The above can be reformulated as follows

$$\mathbf{s} = \mathbf{g} - \frac{\mathbf{w} \mathbf{g} \cdot \mathbf{w}}{\|\mathbf{w}\| \|\mathbf{w}\|} = \mathbf{g} - \frac{\mathbf{w} \mathbf{g}^T \mathbf{w}}{\mathbf{w}^T \mathbf{w}}. \tag{2}$$

With the above formulations, the error function can be stated as follows

$$e(\mathbf{x}, \mathbf{x}_0) = \|\mathbf{s}\| = \|\mathbf{g}(\mathbf{x}) - \mathbf{w}(\mathbf{x}_0) \mathbf{g}^T(\mathbf{x}) \mathbf{w}(\mathbf{x}_0)\|. \tag{3}$$

The total error Ξ is obtained by integrating the square of (3) over all possible locations \mathbf{x} in the neighbourhood Ω , using a Gaussian soft averaging filter G_σ (Brox et al. 2006). That is

$$\begin{aligned} \Xi &= \int_{\Omega} \|\mathbf{g}(\mathbf{x}) - \mathbf{w}(\mathbf{x}_0) \mathbf{g}^T(\mathbf{x}) \mathbf{w}(\mathbf{x}_0)\|^2 G_\sigma(\mathbf{x}, \mathbf{x}_0) d\mathbf{x} \\ &= \int_{\Omega} (\mathbf{g}^T \mathbf{g} - 2\mathbf{g}^T \mathbf{g} + \mathbf{w}^T \mathbf{g} \mathbf{g}^T \mathbf{w}) G_\sigma(\mathbf{x}, \mathbf{x}_0) d\mathbf{x} \\ &= - \int_{\Omega} \mathbf{g}^T \mathbf{g} G_\sigma(\mathbf{x}, \mathbf{x}_0) d\mathbf{x} + \mathbf{w}^T \left(\int_{\Omega} \mathbf{g} \mathbf{g}^T G_\sigma(\mathbf{x}, \mathbf{x}_0) d\mathbf{x} \right) \mathbf{w}. \end{aligned} \tag{4}$$

Now, to find \mathbf{w} we have to solve the following optimization problem

$$\min_{\mathbf{w}} \|\Xi\|, \tag{5}$$

subject to the constraint $\mathbf{w}^T \mathbf{w} = 1$. Solving this problem is equivalent to the maximization problem

$$\min_{\mathbf{w}} \|\Xi\| = \max_{\mathbf{w}} \left\{ \mathbf{w}^T \left(\int_{\Omega} \mathbf{g} \mathbf{g}^T G_\sigma(\mathbf{x}, \mathbf{x}_0) d\mathbf{x} \right) \mathbf{w} \right\}_{\mathbf{w}^T \mathbf{w} = 1}, \tag{6}$$

where the expression

$$\mathbf{T} = \int_{\Omega} \mathbf{g} \mathbf{g}^T G_\sigma(\mathbf{x}, \mathbf{x}_0) d\mathbf{x}, \tag{7}$$

denotes the structural tensor. Using the method of Lagrange multipliers the best orientation \mathbf{w} can be computed as an eigenvector of \mathbf{T} corresponding to its largest eigenvalue. It is given as follows:

$$\mathbf{w} = [T_{11} - T_{22}, 2T_{12}]^T, \tag{8}$$

where T_{ij} are components of \mathbf{T} . Very useful feature is the phase φ of \mathbf{w} , given by the following formula

$$\varphi(\mathbf{w}) = \text{atan2}(2T_{12}/(T_{11} - T_{22})), \tag{9}$$

where atan2 is an extended arctangent function (Cyganek 2013). Values of the phase φ are taken to the phase histograms used in one of the classifiers of our system. Nevertheless, the vector \mathbf{w} is well defined only in the areas with strong structure. This values is indicated by the coherence component c which can be expressed with help of the eigenvalues $\lambda_{1,2}$ of \mathbf{T} , as follows

$$\begin{aligned} c &= ((\lambda_1 - \lambda_2)/(\lambda_1 + \lambda_2))^2 \\ &= \|\mathbf{w}\|^2 / (T_{11} + T_{22})^2, \text{ for } \lambda_1 + \lambda_2 \neq 0. \end{aligned} \tag{10}$$

Values of c which are close to 1 are characteristic of the places with very strong local orientation, whereas values near 0 indicate either anisotropic structure or no signal changes. The latter two conditions can be easily verified analyzing the cumulative magnitude, given as follows

$$m = T_{11} + T_{22}. \tag{11}$$

Thus, values of φ are taken only in and around the areas with high value of m and c . In practice this is achieved setting some threshold τ_m and τ_c . Then, for a given point \mathbf{x} for which $m(\mathbf{x}) \geq \tau_m$ its local neighbourhood $A(\mathbf{x})$ is also taken for computations, if it holds that

$$\forall i : \{\mathbf{x}_i \in A(\mathbf{x}); c(\mathbf{x}_i) \geq \tau_c\}, \tag{12}$$

where c is given in (10). In our experiments the best results were obtained with $\tau_m \approx 0.05$ and $\tau_c \approx 0.5$, whereas m and c were normalized to the range $[0, 1]$. The regions $A(\mathbf{x})$ are squares of 2–15 pixels side. Once computed, the eigenvalues $\lambda_{1,2}$ are further used for computations of corner points, as will be presented. The histograms of orientation phases (9) are used to compare global shapes. However, this method is augmented by computation of the morphological scale-space, as will be discussed.

Figure 4 depicts examples of the structural tensor components: T_{xx}, T_{xy}, T_{yy} , as well as coherence c , phase φ , and magnitude m , computed for one of the test logo patterns belonging to the logo database prepared by the University of Maryland (University of Maryland 2013; Doermann et al. 1996; Zhu and Doermann 2007).

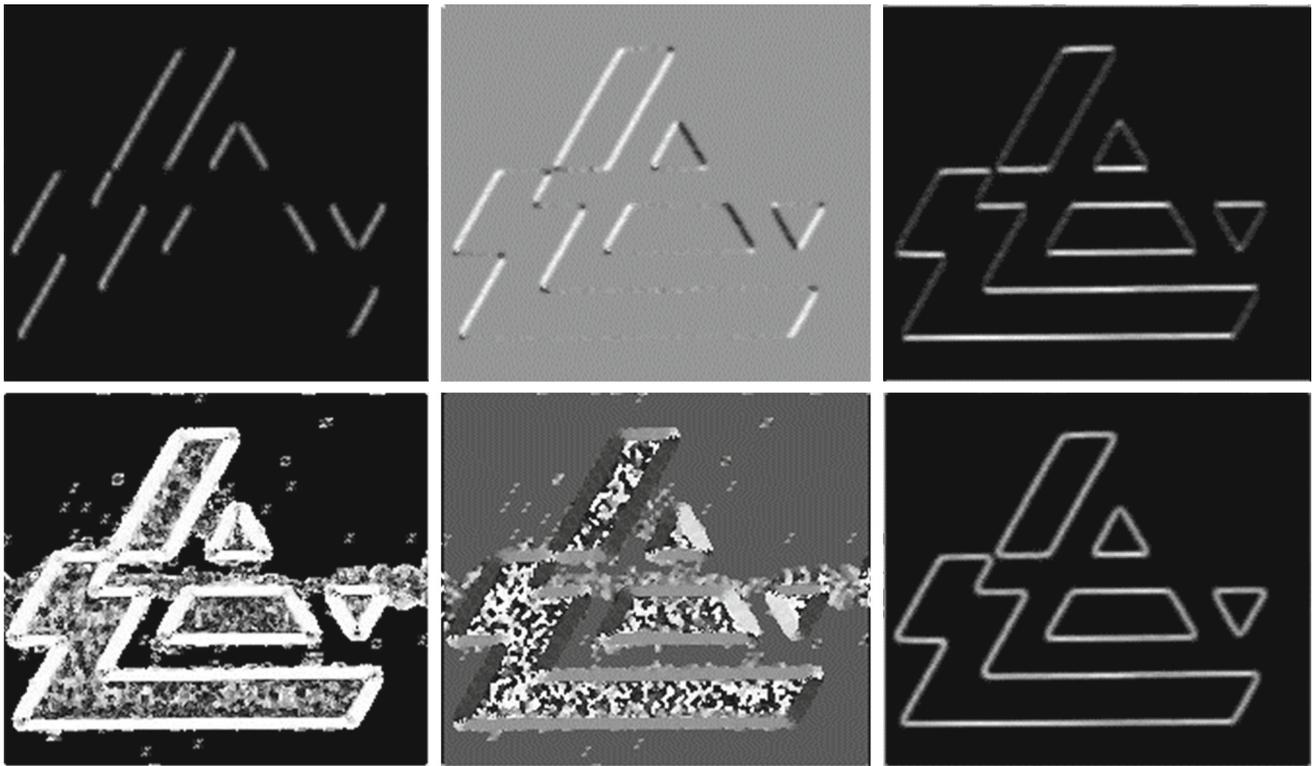


Fig. 4 Examples of the structural tensor components T_{xx}, T_{xy}, T_{yy} , (upper row) coherence, angle, and magnitude (lower row) computed for one of the logo patterns from the database of the University of Maryland

Morphological scale-space

To take advantage of the whole contents of the pictograms contained in the logo images, it is proposed to perform a number of morphological erosions and dilations, each followed by computation of the ST. This way the morphological scale-space (MSS) is obtained. For a signal $f(\mathbf{x})$ the MSS is defined as follows: (Jackway and Deriche 1996)

$$m(f, s_\sigma) = \begin{cases} d(f(\mathbf{x}), s_\sigma) & \text{if } \sigma > 0 \\ f(\mathbf{x}) & \text{if } \sigma = 0 \\ e(f(\mathbf{x}), s_\sigma) & \text{if } \sigma < 0 \end{cases}, \quad (13)$$

where σ denotes scale of the space, $d(f(\mathbf{x}), s_\sigma)$ and $e(f(\mathbf{x}), s_\sigma)$ are the morphological dilation and erosion operators, respectively. As pointed out by Jackway and Deriche (1996), the structuring element has to comply also with additional conditions to avoid unwanted shifts of the features in the processed image. These are fulfilled by the following zero-shift structuring element

$$\sup_{y \in S} [s(\mathbf{y})] = 0 \quad \text{and} \quad s(0) = 0. \quad (14)$$

In the above we have a notion of a positive scale ($\sigma > 0$), which corresponds to the morphological dilation, whereas for negative scales ($\sigma < 0$) erosion is assumed. In a consequence, positive scales correspond to the local maxima of

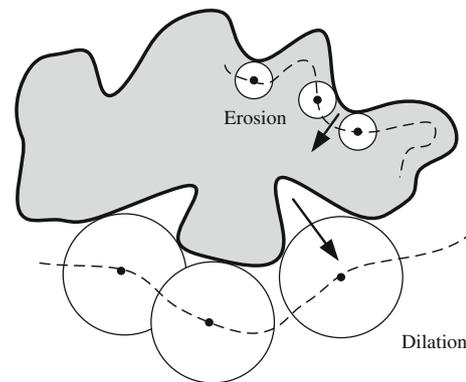


Fig. 5 Example of the morphological scale-space computed from a shape. The shape is carved by a number of erosions and dilations. This way obtained shapes are used to compute histograms of the local phase of the structural tensor

an image, while negative to the local minima. Obviously, for larger modules of σ the filtered image contains less details. Figure 5 depicts a process of creating of the MSS over a grayed shape.

Further it can be shown that the MSS operator $m(f, s_\sigma)$ preserves number and positions of each of the image extreme up to a point called the intrinsic scale. This allows the MSS to build pyramids in an analogous way as the linear scale space pyramids (Cyganek 2013). Thus, the MSS is suitable

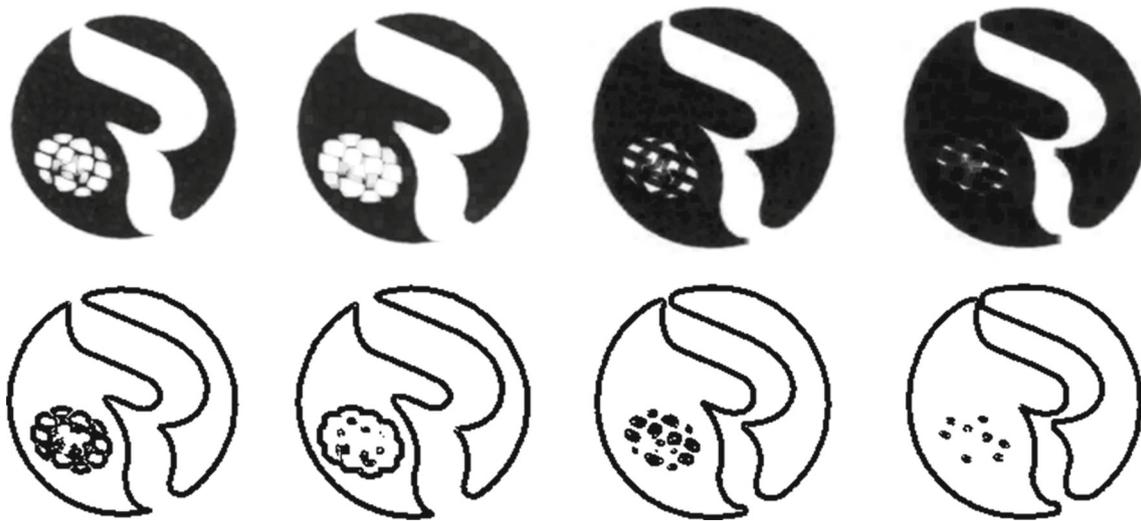


Fig. 6 Two dilated and two eroded version of the logo pattern (*top row*). Each level of the morphological scale-space reveals a different level of image details. From each a global histogram of a phase of the structural tensor in highly structural places is computed (*bottom row*)

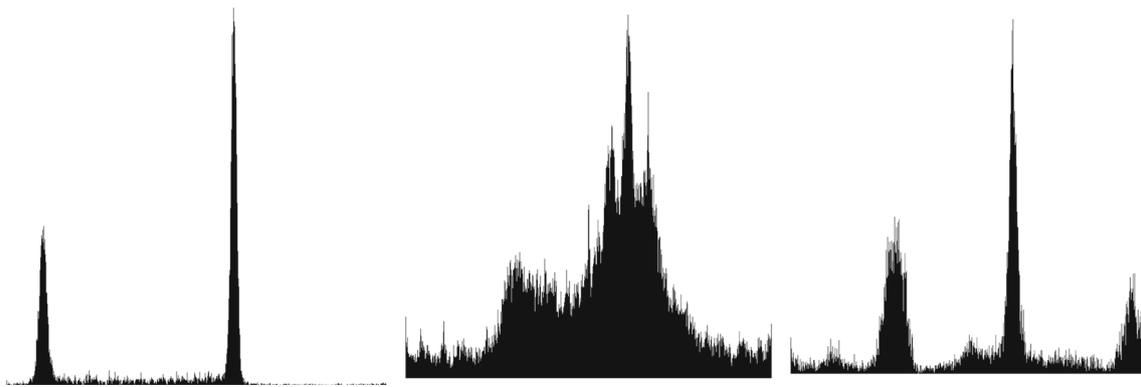


Fig. 7 Exemplary histograms of the local phases of the morphological structural tensor of the logo patterns. Rotations of the pattern are transformed into histogram shifts. Thanks to histogram normalization, also different scales are well accommodated. Number of bins is set from 360 to 720

for coarse-to-fine tracking of features. That is, if a signal feature, which usually corresponds to some form of extreme in an image, appears at a scale $\sigma_i < 0$. Then it also appears at a zero scale and all scales up to a certain intrinsic scale, which depends on a size of a given feature.

Figure 6 depicts two dilated and two eroded version of one of the test logo patterns. Each level of the MSS reveals a different level of image details. From each a global histogram of a phase of the structural tensor in sufficiently structural places is computed. Thanks to the properties of the MSS this will lead to the enhancement of the representation of the characteristic local features. Thanks to this method, phases (orientations) of reappearing features are boosted in the histograms. Exemplary histograms are shown in Fig. 7.

As we verified experimentally, the chosen comparison measure of the histograms leads to very different results. For this purpose the two histogram comparison measures were investigated. The first one is the Bhattacharyya dis-

tance, given as follows (Aherne et al. 1998; Bhattacharayya 1943):

$$D_B(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^{\rho} \sqrt{P(a_i)P(b_i)}. \tag{15}$$

However, the better results were obtained using a version of the following general formula (Cyganek 2013):

$$D_{nm}(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^{\rho} \frac{|P(a_i) - P(b_i)|^n}{[P(a_i) + P(b_i)]^m}. \tag{16}$$

Values of $n = 2$ and $m = 1$ give the χ^2 statistics distance. However, in our experiments better results were obtained for $n = m = 2$. Such measure promotes matches on bins with higher probability. However, in the latter case, the value of D_{nm} in (16) can be greater than 1.

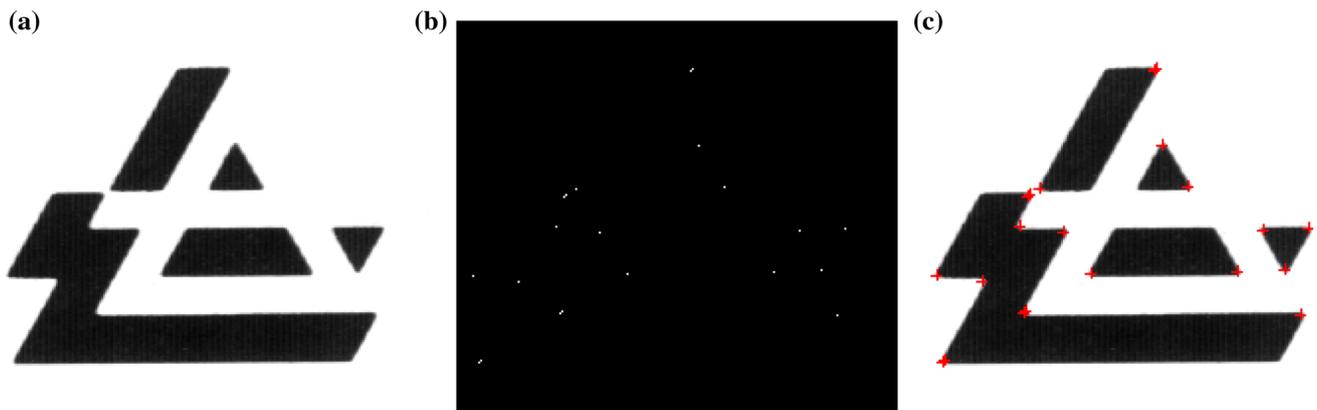


Fig. 8 Examples of found corner points in a logo image. The original test image (a). Corner points in the image space (b), the points superimposed on the pattern (c)

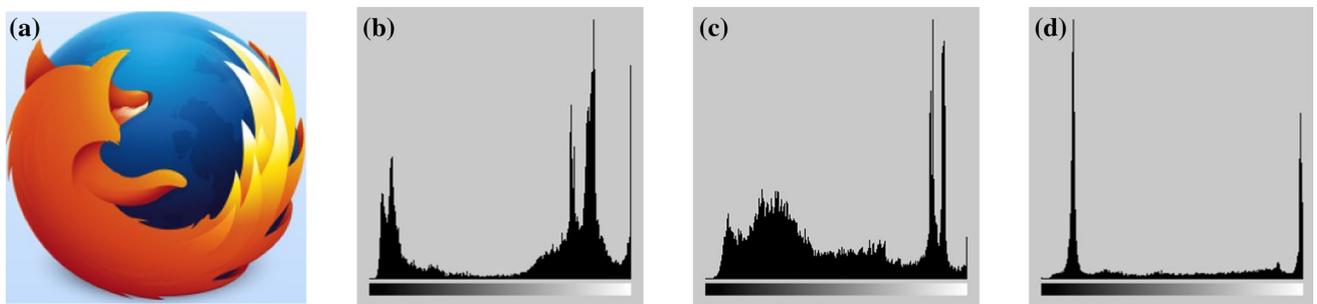


Fig. 9 An example of a color logo (a) and three histograms of its red (b), green (c), and blue (d) color channels

Computation of corner points with the structural tensor

Positions of corner points provide important information on image contents. In the limited case of highly contrasted logo patterns, these can be easily found and used for pattern matching, as will be discussed. After computing the ST, the corner points (x_i, y_i) can be found as the ones which fulfill the following condition

$$\lambda_1(x_i, y_i) \gg \lambda_2(x_i, y_i) > \kappa, \quad (17)$$

where κ is a threshold for the lower eigenvalue of the structural tensor. It can be set to 0 if the point priority technique is employed, as described in Cyganek and Siebert (2009). Introducing (10) into (17), the following condition is obtained:

$$[T_{11}(x_i, y_i) - T_{22}(x_i, y_i)]^2 + 4T_{12}^2(x_i, y_i) > \kappa^*. \quad (18)$$

where κ^* denotes a new threshold.

Figure 8 shows examples of corner detection with the ST in one of the test images. In our framework it is further possible to divide an image into a number of tiles, in which corner points are then detected. Thanks to such strategy, a more uniform distribution of corner points is obtained. However, we observed also that such tiles cannot cope with patterns

which are further rotated or, in general, affinely transformed. Therefore, in the further experiments, the corner points are searched uniformly in the whole image.

4.2 Color discrimination module

Although logo and trademark symbols can be efficiently compared exclusively from based on their shape, in some applications color is also taken into consideration. For this purpose the color discriminative module was designed which can sift out images with different color contents irrespective of their shapes.

The method is based on simple and independent comparison of the normalized color channels histograms in the RGB space. Figure 9 shows an exemplary logo color image, and its three R-G-B color channels, respectively. Although color components are not independent, this simple scheme works well in practice since only coarse selection of input images is necessary at this stage. Also, other color spaces, such as HSI or Lab, can provide better discrimination. Using the RGB color space we save on color space conversion, however. For a test color image, its fit measure can be assessed first independently comparing the color channels histograms. Then,

the color distance between images is computed as follows

$$D_B = \prod_{i=1}^3 D_i. \tag{19}$$

where D_i denotes a match value of the i th histograms, as provided in the formulas (15) and (16). Sensitivity of this comparison can be controlled setting proper size of the bins in the histograms (Cyganek 2013). In the experiments these were set from 16 to 64. For more advanced color discrimination other techniques can be used, such as a method based on the ensemble of one-class SVM classifiers (Cyganek and Wiatr 2011).

4.3 Logo recognition with the affine moment invariants

The affine moment invariants (AMI) were proposed for pattern recognition by Flusser and Suk (1994). They are based on the theory of algebraic moment invariants and are invariant to the general affine transformation which takes a 2D point \mathbf{x} into the corresponding $\hat{\mathbf{x}}$ in accordance with the following formula

$$\hat{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}, \text{ where } \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}. \tag{20}$$

It was shown that the following first four AMI are sufficient for efficient pattern recognition in many vision tasks (Flusser et al. 2009; Flusser and Suk 1994; Cyganek 2013):

$$\begin{aligned} \pi_1 &= c_{00}^{-4}(c_{20}c_{02} - c_{11}^2), \\ \pi_2 &= c_{00}^{-10}(c_{30}^2c_{03}^2 - 6c_{30}c_{21}c_{12}c_{03} + 4c_{30}c_{12}^3 \\ &\quad + 4c_{03}c_{21}^3 - 3c_{21}^2c_{12}^2), \\ \pi_3 &= c_{00}^{-7}[c_{20}(c_{21}c_{03} - c_{12}^2) - c_{11}(c_{30}c_{03} - c_{21}c_{12}) \\ &\quad + c_{02}(c_{30}c_{12} - c_{21}^2)], \\ \pi_4 &= c_{00}^{-11}[c_{20}^3c_{03}^2 - 6c_{20}^2c_{11}c_{12}c_{03} - 6c_{20}^2c_{02}c_{21}c_{03} \\ &\quad + 9c_{20}^2c_{02}c_{12}^2 + 12c_{20}c_{11}^2c_{21}c_{03} \\ &\quad + 6c_{20}c_{11}c_{02}c_{30}c_{03} - 18c_{20}c_{11}c_{02}c_{21}c_{12} \\ &\quad - 8c_{11}^3c_{30}c_{03} - 6c_{20}c_{02}^2c_{30}c_{12} + 9c_{20}c_{02}^2c_{21}^2 \\ &\quad + 12c_{11}^2c_{02}c_{30}c_{12} - 6c_{11}c_{02}^2c_{30}c_{21} + c_{02}^3c_{30}^2], \end{aligned} \tag{21}$$

where c_{ab} are central moments defined as follows

$$c_{ab} = \sum_{r=1}^R \sum_{s=1}^S I_{rs}(r - \bar{x})^a(s - \bar{y})^b. \tag{22}$$

In the above formulas I_{rs} denotes scalar image intensity value at discrete position (r, s) . The one problem with pattern recognition with AMI features lies in comparing AMI values which frequently have highly different exponents. For this purpose we propose to factor each value of the AMI values, i.e. those expressed in (21), into mantissa and exponent. Then, when testing the incoming test AMI values, these are first scaled in concordance with the corresponding exponent

computed in the training patterns. Then, the obtained values, which are now free from the high dynamic range, are normalized and compared with the Euclidean distance. A perfect match gives 0, whereas a total mismatch leads to 1.

4.4 Hausdorff log-polar augmented distance for point matching

Point correspondence is a well known technique of image matching (Cyganek and Siebert 2009). In this approach, intensity values or other descriptors around the characteristic points are compared. However, geometrical configurations of the points are not considered. On the other hand, the Hausdorff distance allows geometrical comparison of point sets, not considering intensities of the contained points. In this paper we propose composition of the two approaches to take benefit of both. The idea of joining these two domains is not a new one. For instance, Yang et al. (2007) propose to join the Hausdorff distance with the normalized gradient matching approach. However, in this paper we propose to join the Hausdorff measure with the log-polar (LP) image matching of image patches around the corner points. The log-polar transform allows effective matching even for the rotated and scaled images. This is achieved since rotation and scale change in the Cartesian coordinate system translates into horizontal and vertical shifts in the log-polar domain (Zokai and Wolberg 2005; Cyganek 2013). The image patches are considered only around the salient corner points found alongside with the computation of the structural tensor, as already discussed. Contrary to edge points, the number of corner points is highly limited which facilitates computations. Finally, the geometrical configuration of the matched point sets is taken into consideration thanks to the properties of the Hausdorff distance. This novel combination showed very good properties when comparing images, as will be discussed.

The Hausdorff distance, introduced into computer vision by Huttenlocher et al. (1993), finds the maximal inter-pixel distance between the two point sets $A = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{N_A}\}$ and $B = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_{N_B}\}$, each containing N_A and N_B points, respectively. The classical definition of the Hausdorff distance is as follows

$$D_H(A, B) = \max_{m,n} (d_S(\mathbf{a}_m, B), d_S(\mathbf{b}_n, A)), \tag{23}$$

where $d_S(\mathbf{a}_m, B)$ denotes a distance from a point \mathbf{a}_m to a set of points B . It is defined as

$$\begin{aligned} d_S(\mathbf{a}_m, B) &= \min_{\mathbf{b}_n \in B} (d(\mathbf{a}_m, \mathbf{b}_n)), \text{ and similarly} \\ d_S(\mathbf{b}_n, A) &= \min_{\mathbf{a}_m \in A} (d(\mathbf{b}_n, \mathbf{a}_m)), \end{aligned} \tag{24}$$

where $d(\mathbf{a}_m, \mathbf{b}_n)$ in turn, is a defined distance between the points \mathbf{a}_m and \mathbf{b}_n . The distance $d(\mathbf{a}_m, \mathbf{b}_n)$ can be defined with any metric, such as the city-block or the chessboard ones.

However, in the presented system we use only the Euclidean distance, defined as follows

$$d(\mathbf{a}, \mathbf{b}) = \left(\sum_{k=1}^L (a_k - b_k)^2 \right)^{1/2}, \quad (25)$$

where a_k and b_k denote k th coefficient of the points \mathbf{a} and \mathbf{b} , respectively, and L is dimension of the space ($L = 2$ for 2D images).

The above classical Hausdorff definition has been extended by many authors to suit better in the domain of pattern recognition. For instance, [Dubuisson and Jain \(1994\)](#) define a directed distance from a set of points A to the set B , as follows

$$d_1(A, B) = \min_{\mathbf{a}_m \in A} (d_S(\mathbf{a}_m, B)), \quad (26)$$

$$d_2(A, B) = \max_{\mathbf{a}_m \in A} (d_S(\mathbf{a}_m, B)), \quad (27)$$

$$d_3(A, B) = \frac{1}{N_A} \sum_{\mathbf{a}_m \in A} d_S(\mathbf{a}_m, B), \quad (28)$$

$$d_4(A, B) = Q(q, d_S(\mathbf{a}_m, B)), \text{ for } \mathbf{a}_m \in A. \quad (29)$$

In the above, Q denotes a q th quantile, and for $q = 0.5$ it is a median. Based on the above directed set-to-set distances, the following undirected distances are further defined, as follows

$$U_1(A, B)|_i = \min(d_i(A, B), d_i(B, A)), \quad (30)$$

$$U_2(A, B)|_i = \max(d_i(A, B), d_i(B, A)), \quad (31)$$

$$U_3(A, B)|_i = \frac{1}{2}(d_i(A, B), d_i(B, A)), \quad (32)$$

$$U_4(A, B)|_i = \frac{N_A d_i(A, B) + N_B d_i(B, A)}{N_A + N_B}. \quad (33)$$

It is easy to observe that the classical Hausdorff distance D_H in (23) is obtained for U_2 setting $i = 2$.

The above distances have been also used by [Yang et al. \(2007\)](#) in their system for hybrid image matching. However, Yang et al. also proposed to extend the above distances adding a kind of a third dimension which conveys information on distance of the image patches around the points \mathbf{a} and \mathbf{b} . This patch related distance can be simply sum of absolute differences SAD between the corresponding pixels or the normalized gradient ([Cyganek and Siebert 2009](#)). However, this additional information cannot cope with high variations of image patch rotations and scale changes. Therefore in this paper we propose to use the log-polar extended distance which can well accommodate rotation and scale changes between the image patches. Such defined Hausdorff distance will be called undirected log-polar Hausdorff distance (ULPH). Such LP patch matching was first proposed by [Zokai and Wolberg \(2005\)](#) for image registration, and then also used in the sparse stereo image matching which details are presented in [Cyganek and Siebert \(2009\)](#). Briefly, the log-polar transformation of a point \mathbf{a} from the Euclidean space

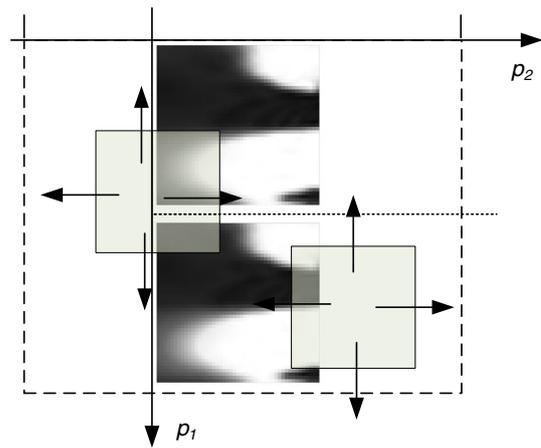


Fig. 10 Search space for image patch matching in the log-polar space. Since rotation and scale in Euclidean space translates into vertical and horizontal shifts in the log-polar space, the reference pattern is warped and the best fit value and position reported

into the point \mathbf{p} in the LP space is defined as follows

$$\begin{aligned} p_1 &= \log_B(\sqrt{(a_1 - c_1)^2 + (a_2 - c_2)^2}), \\ p_2 &= \text{atan2}\left(\frac{a_2 - c_2}{a_1 - c_1}\right) \end{aligned} \quad (34)$$

where $\mathbf{c} = (c_1, c_2)$ is a central point of the LP transformation, B denotes a logarithm base. In our system the central point \mathbf{c} for each patch of an image is a corner point found during computation of the structural tensor, as already discussed in [Sect. 4.1](#).

As already mentioned, the Hausdorff distance conveys information on geometrical similarity of the two sets of points. However, in real applications this measure can be augmented taking into consideration correspondence values measured for each pair of points. Patch fitting in the LP space requires 4D match in the warped pattern space, as depicted in [Fig. 10](#). Rotation and scale in Euclidean space translates into vertical and horizontal shifts in the log-polar space. Therefore the reference pattern is warped and the best fit value and position reported as a match result. This process is depicted in [Fig. 11](#). These values are added to the Hausdorff distance between point sets in the two matched images. Further details on this method can be found in [Cyganek \(2013\)](#); [Zokai and Wolberg \(2005\)](#).

Thus, the proposed point-to-point distance (25) is augmented with the LP matched values is defined as follows

$$D_{H-LP}(\mathbf{a}, \mathbf{b})|_i = w_1 U_i(\mathbf{a}, \mathbf{b}) + w_2 D_{LP}(\mathbf{a}, \mathbf{b}), \quad (35)$$

where $D_{LP}(\mathbf{a}, \mathbf{b})$ denotes a match between the image patches around the points \mathbf{a} and \mathbf{b} , respectively, w_1 and w_2 are weights defining influence of each component in the final distance. In our experiments $w_1 = 1$ and $w_2 = 100$ since the D_{LP}

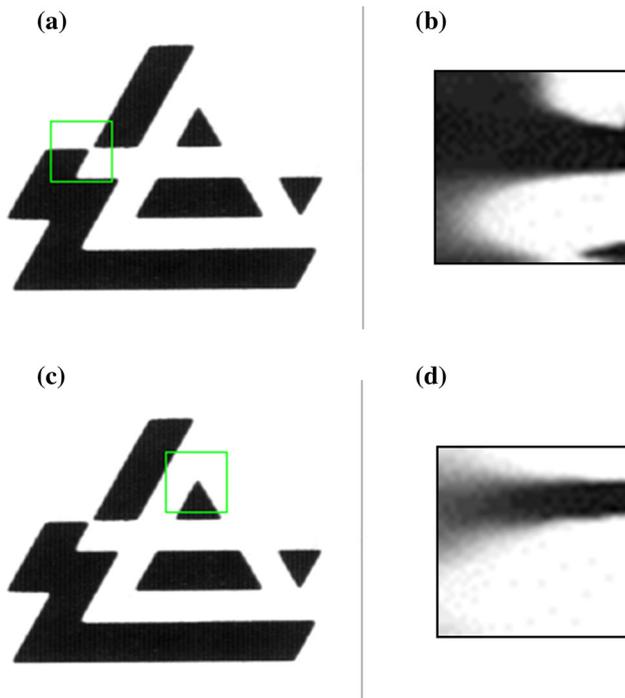


Fig. 11 Exemplary logo image with patches (green rectangles) around corner points (a–c). Log-polar representation of the local image patches (b–d)

measure is from 0 to 1. For the D_{LP} measure the covariance-variance distance was used.

$$D_{LP}(\mathbf{a}, \mathbf{b}) = \frac{\sum_S (I_1(\mathbf{a}) - \overline{I_1(\mathbf{a})}) \cdot (I_2(\mathbf{b}) - \overline{I_2(\mathbf{b})})}{\sqrt{\sum_S (I_1(\mathbf{a}) - \overline{I_1(\mathbf{a})})^2 \cdot \sum_S (I_2(\mathbf{b}) - \overline{I_2(\mathbf{b})})^2}}, \tag{36}$$

where $I_1(\mathbf{a})$ and $I_2(\mathbf{b})$ are two log-polar image patches around the points \mathbf{a} and \mathbf{b} , respectively, $\overline{I_1(\mathbf{a})}$ and $\overline{I_2(\mathbf{b})}$ are average values of each patch, respectively, expressed in the log-polar space and S denotes the set of point coordinates belonging to image patches. The drawback of the measure (36) is its long computation time. Therefore other faster measures were also checked, such as Census based or SAD. Nevertheless, the best results were obtained with (36) since it is invariant to intensity changes. For this purpose the GPU implementation of the LP patch matching is implemented which allows real-time operation.

5 Combining classification modules into a hybrid ensemble

The classifiers are joined into an ensemble which usually shows much better final accuracy on the whole set of test patterns when compared to each of the single classifiers when

operating alone. However, the classifiers can be joined in many different ways, as already discussed. The two most characteristic are the serial versus the parallel arrangements. Usually, the hybrid serial-parallel compositions operate best in practice. In our system, the optional color classifier was set as the first one, whereas all other were operating in a parallel fashion. This means that results output by the first classifier are fed to the group of parallel classifiers. The architecture is presented in Fig. 12.

When composing an ensemble of classifiers operating on the same level of a parallel branch, the final answer is given based on the chosen fusion method. However, output values of some of the classifiers do not lie in the range [0,1]. Therefore they need to be converted, so the outputs of base classifiers can be joined together. For such outputs which are not in the range [0,1] the following function is used

$$f(x) = \frac{1}{1+x}, \tag{37}$$

where x is a positive match measure. This function for $x = 0$ gives the maximal value 1, whereas for $x > 0$, $f(x)$ decays toward 0.

For fusion of the numerical outputs of the base classifiers, in the proposed system the following approaches were examined (Kuncheva 2004; Polikar 2006; Woźniak 2013):

- A product rule

$$\mu(x) = \prod_{c=1}^C D_c(x), \tag{38}$$

- A sum rule

$$\mu(x) = \frac{1}{C} \sum_{c=1}^C D_c(x), \tag{39}$$

where D_c denotes an output of the c th classifier in the ensemble. In experiments the first rule showed to provide better results. Therefore all further experimental results assume the product rule (38).

Proper composition of the ensemble can be seen as selection of its right architecture for a given recognition problem, as well as settings of their parameters. This can be seen in terms of expert knowledge, and can be expressed in terms of the IF-THEN rules for a given class of logo recognition problems.

Last but not least, apart from the higher accuracy of the ensemble when compared with performance of each single classifier, due to the parallel organization of the proposed hybrid ensemble, operation of each member classifier can be also done in parallel. i.e. by a separate thread. Only operation of the fusion module needs to be preceded by synchro-

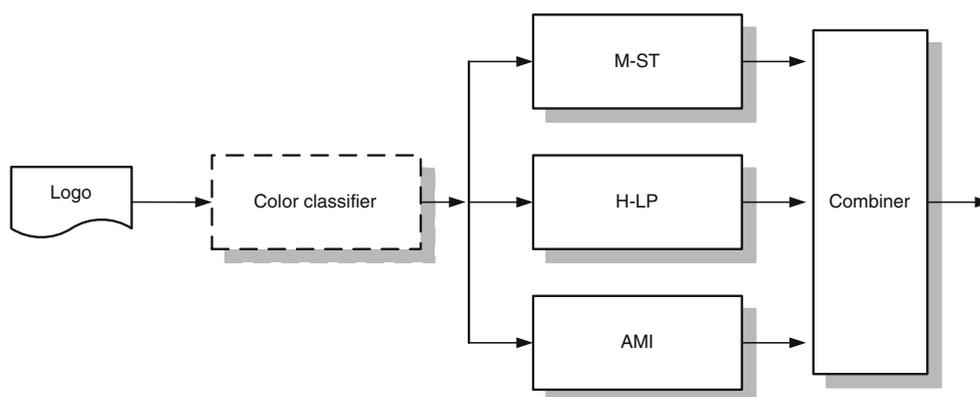


Fig. 12 Block diagram of the ensemble of classifiers. The output of the color classifier is fed into a group of classifiers operating in parallel: the Morphological Structural Tensor (M-ST), the Hausdorff Log-Polar

(H-LP), and the Affine Moment Invariants (AMI). The combiner realizes the product rule. Outputs of each of the classifiers in the ensemble are conditioned to fit into the range $[0,1]$

nization of the member classifiers. This greatly facilitates computations and allows real-time operation due to proper architecture, as will be discussed in the experimental section.

6 Experimental results

The presented method was implemented in C++ using the *HIL* and *DeRecLib* software libraries (Cyganek 2013; Cyganek and Siebert 2009). Experiments were run on the PC with 8 GB RAM and Pentium Q 820 multi-core micro-processor. The member classifiers were implemented with help of the composite design pattern which allows abstraction on building hierarchical structures of the contained objects (Cyganek and Siebert 2009). For the experiments the University of Maryland Logo Dataset from the Laboratory for Language and Media Processing (LAMP) of the University of Maryland was used (University of Maryland 2013; Doermann et al. 1996; Zhu and Doermann 2007). The database contains 105 logo images which, for our experiments, were affinely converted to create the test patterns. Exemplary images from this database are shown in Fig. 13.

In experiments the two scenarios were investigated:

- Accuracy of recognition of the affinely deformed patterns from the test database for different compositions of the hybrid ensemble;
- Generalizing abilities of the ensemble of classifiers;

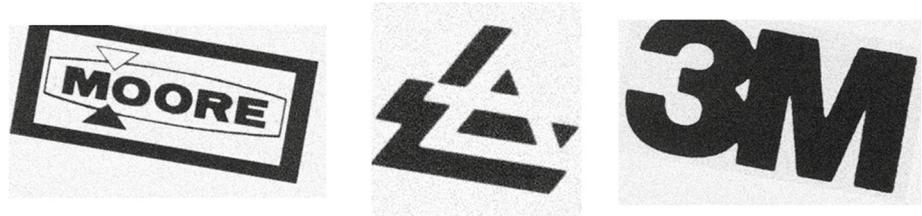
The first group of tests shows recognition abilities of the proposed ensemble of classifiers. In other words, the answer how resistant the system is to rotation, change of scale, occlusions, as well as additive noise of the prototype patterns is examined.



Fig. 13 Selected patterns from the University of Maryland database used in the experiments (University of Maryland 2013; Doermann et al. 1996; Zhu and Doermann 2007)

In the paper by Yang et al. (2007) only 20 images from the database were investigated. However, in our experiments all images from the University of Maryland were used. The tests were performed to check performance of the method in respect to different deformations of the database. More precisely, the rotated from -35 up to $+35$ degrees, as well as x and y scale scaled versions, in the range 0.7 up to 1.3, were generated. Then extensive tests, which consisted in comparing each original image from the database with each deformed version, were run. In other words, each orig-

Fig. 14 Three patterns with added Gaussian noise at level of 10%



inal logo image is compared with 105 test images. However, even more interesting is to investigate a returned set of the most similar images to a given one in order to assess the generalization properties of the proposed method. As already mentioned, the proposed framework allows different configurations of the ensemble.

To test influence of noise on classification, the Gaussian noise was added to some of the test patterns. Adding Gaussian noise to the image intensity signal $I(x, y)$ can be expressed as follows

$$\hat{I}(x, y) = I(x, y) + a\eta, \quad (40)$$

where η denotes a random variable of the Gaussian distribution and a is a control parameter. It can be shown that for 8-bit scalar valued images and normal distribution, i.e. zero mean and $\sigma = 1$ the Peak-Signal to Noise Ratio (PSNR) can be expressed as follows (Cyganek and Siebert 2009)

$$\text{PSNR}(\hat{I}, I) = 48.13 - 20 \log_{10}(a), \quad (41)$$

In the above, the normal random generator is assumed and the parameter a controls influence on original signal. Thus, a can be set in a relation to the maximal pixel value (i.e. 255) and expressed as a percentage of the signal amplitude. In other words, $a = 0\%$ represents a pure intensity, while $a = 100\%$ leads to $\text{PSNR} = 0$. In our experiments, the noise affected test patterns were generated with the parameter a varying from 5 to 50%. Examples of some test patterns, contaminated with the Gaussian noise, are presented in Fig. 14.

Table 1 contains description of the control parameters of the three member classifiers used in the presented ensemble: M-ST, H-LP, AMI as well as the color pre-selector. Also, default values, i.e. the ones which show the best results, are shown in the last column of Table 1.

All of the experiments of the first group were done to check ability of the system to find each of the logo patterns in respect to the set of all of the logo patterns but transformed by rotation, change of scale, as well as addition of noise. After presenting a test pattern a ranking list of the best answers for that pattern, given by an ensemble, is stored. Then, accuracy is measured assuming the pattern was found on the first, second, and up to the sixth position on the list. In the following plots this is called an answer tolerance. In other words, if the answer tolerance is 1, this means that the correct pattern was

recognized as the second best pattern on the ranking list, and so on.

Figure 16a shows plots of accuracy of single classifiers (black curves) tested on the logo images rotated by 20° which also crops some of the images, as shown in the lower row of Fig. 15. Shown are accuracy curves for the AMI, M-ST, H-LP, as well as their combinations in the ensembles (color curves) E1 (M-St and H-LP) and E2 (M-St, H-LP, and AMI). It is evident that a combination of classifiers in a hybrid ensemble significantly outperforms each single classifier.

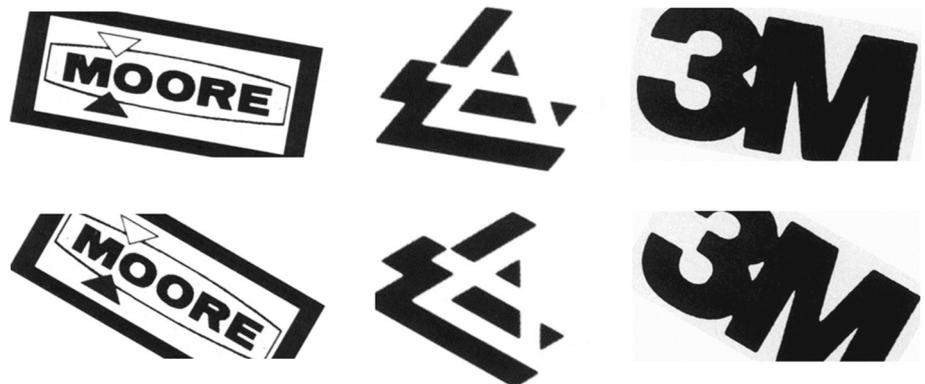
On the other hand, the plot in Fig. 16b shows comparison of differences in the fusion method of the member classifiers, as discussed in Sect. 5. The sum combination method, given in (38), versus the product rule, given in (39), were investigated. The latter showed the best results. Intuitively, the product rule is more sensitive in the cases of one classifier voting against a given test pattern. In other words, for the product rule (38), a single classifier in an ensemble returning low match value affects more the final score than in the case of the sum output rule (39). Nevertheless, it is interesting to observe that the difference in the responses of the product and sum rules is the highest when an exact logo match is requested, i.e. for the 0 answer tolerance. However, if it is acceptable to have a correct match on a certain further position on the ranking list, then the results tend to give similar results, as shown in Fig. 16b. This behavior can be explained by the different competence regions of all of the member classifiers. Indeed, all the member classifiers have different structures and properties, and all operate on different data drawn from the images (such as edges vs. log-polar regions around certain points).

Figure 17 shows an interesting case of the pattern which showed to be very difficult for the M-ST and H-LP classifiers, which for other patterns perform quite well. On the other hand, the AMI classifier does perfect recognition of that one. This is due to the specific properties of that shape which is compact and filled. Nevertheless, this example shows the power of diversity of an ensemble of classifiers in which diversity of the member classifiers leads to much higher overall accuracy of such ensemble.

Figure 18a depicts measured accuracy of the ensemble (M-ST, H-LP, AMI) for different values of rotations 10° – 30° but without change of the contained patterns. For this purpose, size of each of the rotated patterns was accommodated

Table 1 Comparison of control parameters of the classifiers used in the tested ensembles

Type of a classifier	Parameter	Values range	Default values
M-ST	Number of erosions	1–5	2
	Number of dilations	1–5	2
	Differentiating filter type	Sobel	Simoncelli
		Simoncelli	
	Bins in the histogram	360–720	720
Histogram comparison measure	Bhattacharyya (15) χ^2 (16) Modified χ^2	Modified χ^2	
H-LP	Number of corner points per tile	1–33	13
	Number of tiles for corners	1×1 – 10×10	1×1
	Size of the log-polar patch	23×23 – 55×55	23×23
	LP patch comparison measure	CoVar	CoVar
		Census	
		SAD	
	Point set-to-set distance	(26)	(29)
		(27)	
		(28)	
		(29)	
Hausdorff distance	(30)	(31)	
	(31)		
	(32)		
	(33)		
AMI	–	–	–
Color	Color space	RGB	RGB
		HSI	
		Lab	
	Number of histogram bins	16–256	64
Histogram comparison measure	Bhattacharyya χ^2 Modified χ^2	χ^2	

Fig. 15 Three exemplary patterns from the database. Rotated 10° (top), rotated 30° (bottom). Higher rotation angles result in pattern cropping

to fit the rotated pattern. However, higher rotations while keeping the original size of images result in pattern cropping, which makes recognition much harder. This effect is well vis-

ible for higher values of rotation, which leads to a much lower accuracy, as shown in Fig. 18b. Nevertheless, for the tolerance higher than 4 it again reaches high acceptable value.

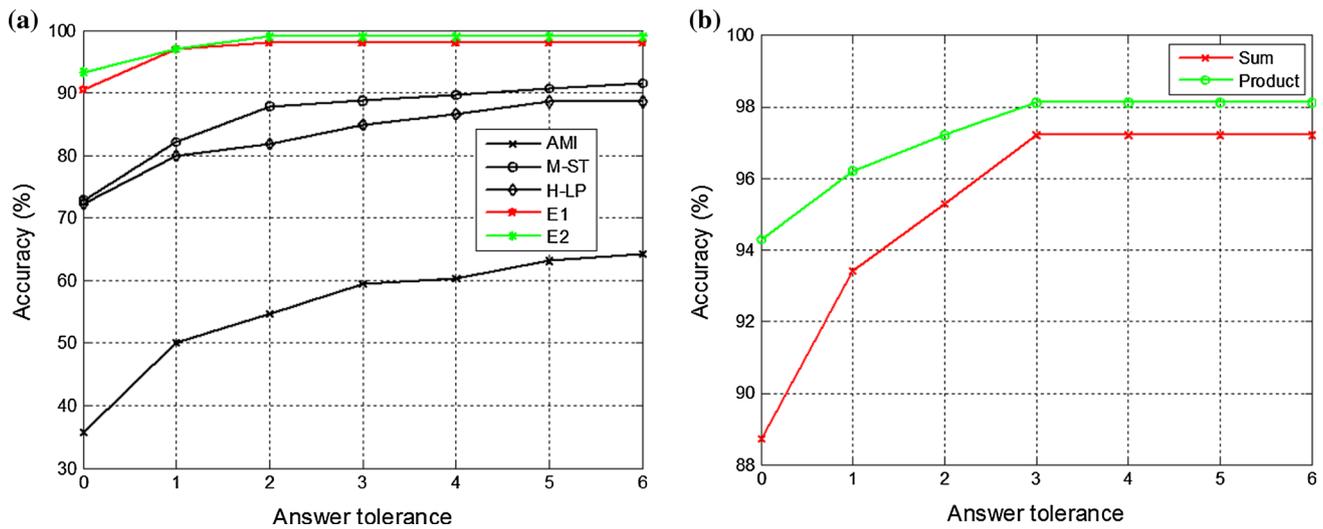


Fig. 16 Accuracy of single classifiers (black curves) performing on 20° rotated and occluded (cropped) test logo patterns: AMI, M-ST, H-LP, as well as their combinations in the ensembles (color curves) E1 (M-St and H-LP) and E2 (M-St, H-LP, and AMI). It is evident that a combination of classifiers in a hybrid ensemble greatly outperforms each

Fig. 17 An example of a logo pattern resulting in high classification errors for the M-ST and H-LP classifiers. However, the AMI based classifiers alone allows its perfect recognition



Figure 19a contains plots of accuracy of the ensemble (M-ST, H-LP, AMI) for 10° rotated patterns in respect to the answer tolerance and different Gaussian noise level (average values shown). In Fig. 19b accuracy of the same ensemble is shown, measured for patterns changed in scale. For scale set to 0.7, the patterns get smaller and thanks to this are perfectly recognized by the ensemble. However, for the higher scale, set to 1.3 in the second plot, this change leads to cropped shapes which results also with lower accuracy, as shown in Fig. 19b.

From the properties of the member classifiers, as well as based on the previous research, we can conclude the following properties of the member classifiers:

1. The AMI classifier works well for recognition of the global shape of patterns. However, it is very sensitive to occlusions, noise and other outliers.
2. The M-ST operates well for rotated and scale changed patterns, accommodating well to shape variations. However, since ST is computed differentiating the input signal, the extensive noise can cause problems. Nevertheless, the method is insensitive to small variations of the patterns.

single classifier (a). Plot comparing differences in the fusion method of the member classifiers: sum vs. product rule which shows better accuracy (b). Answer tolerance denotes a position of the tested pattern in the response list

3. The H-LP takes into consideration the geometrical positions of the corner points, as well as matching value of the surrounding image patches. This sparse method can deal well with moderate pattern occlusions. However, the method relies on proper detection of the salient points. In our case, these are corner points found when computing the ST from the original image.

At a first glance, the above classifiers perform poor if applied alone. Especially the AMI classifier gives unacceptable results when recognizing highly deformed or noisy patterns. However, the situation is spectacularly improved when these weak classifiers are joined together into a cooperating ensemble of a structure depicted in Fig. 12. In such architecture the ensemble performs much better than each of the member classifiers alone. The results are shown in Fig. 16a. It is evident that both ensembles perform 20–30 % better on average than each single member classifier. In this respect the product rule of fusing the output classifiers was found to perform better than the sum rule, as shown in Fig. 16b. This is why ensembles of classifiers can be seen as emerging trends in development of modern classification systems. One of the main issues in such systems is assurance of a high divergence of the member classifiers. This can be done in many ways, for instance providing different datasets to the specializing member classifiers (Krawczyk et al. 2014), or providing classifiers of different properties (Kuncheva 2004; Woźniak 2013). As demonstrated, the system shows high robustness and compares favorably to other reported systems in the domain of logo and trademark recognition (Chen et al. 2003).

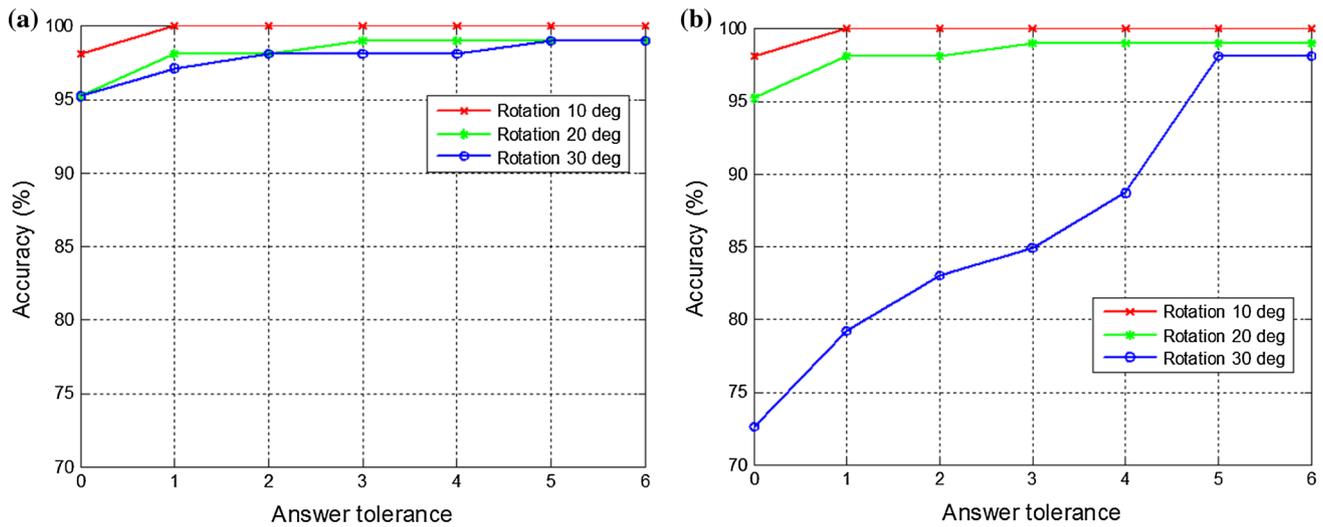


Fig. 18 Accuracy of the ensemble (M-ST, H-LP, AMI) for different values of rotation (a), as well as rotation and occlusion in respect to the answer tolerance (b). The highest occlusion exists for 30° rotation

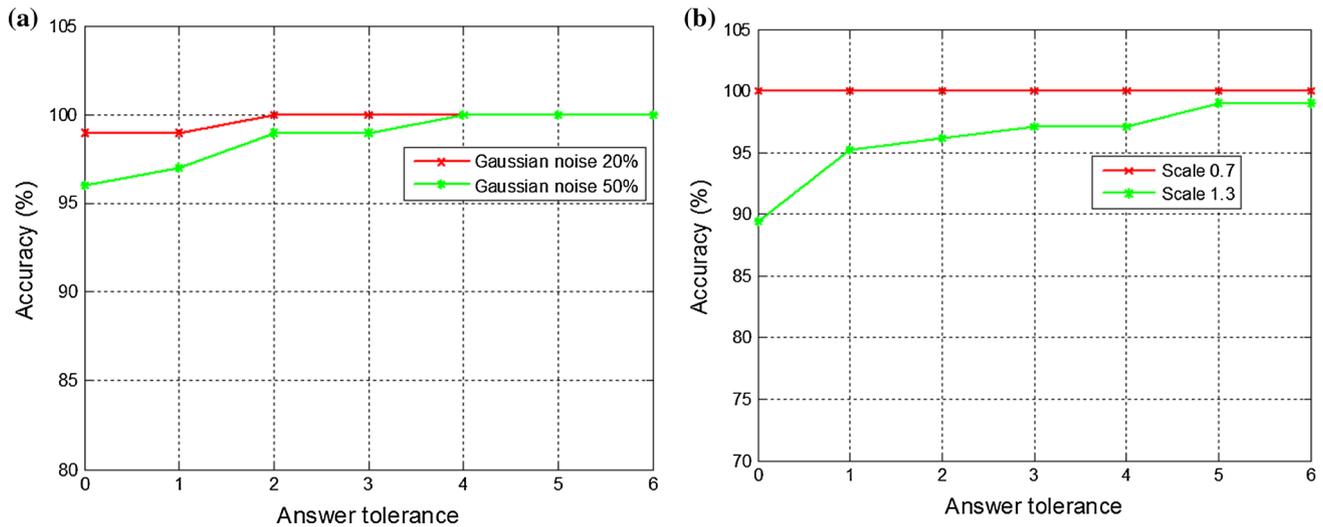


Fig. 19 Accuracy of the ensemble (M-ST, H-LP, AMI) for 10° rotated patterns in respect to the answer tolerance and different Gaussian noise level (a). Accuracy of the same ensemble measured for two scale

changed patterns. Smaller patterns are perfectly recognized (scale 0.7). Higher scale (scale 1.3) leads to changed shape which results with lower accuracy (b)

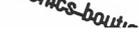
As already mentioned, the second group of experiments was designed to assess the generalization properties of the presented method. In other words, we are interested in answering the question how similar, also in a human like perception, are the returned best matched patterns. Some examples of this experiment, run with the ensemble composed of the two member classifiers M-ST and H-LP and the group of rotated by 20° patterns, are shown in Table 2. Shown images have different dimensions which are only adjusted to fit into the space of Table 2. Each row contains images found

to be most similar to the leftmost one. The values represent responses of the two member classifiers, respectively. These can be used to provide a fuzzy similarity measure between images.

When measuring processing time we noticed that in a serial software implementation the system can process on average 4–5 images per second. However, this grows to 15–20 images with GPU implementation of the LP matching and ST computations. This can be even further parallelized, assigning a separate thread pool for processing of each logo pattern.

Table 2 Similar images in the group of images rotated by 20° found by the combination of the ensemble with the M-ST and H-LP classifiers. Shown logos have different dimensions (here adjusted to fit into the page). Each row contains images found to be most similar to the

leftmost one. The values represent responses of each of the member classifiers. These can be used to provide a fuzzy similarity measure between images

Searched logo	Response of the 1st place	2nd place	3rd place	4th place	5th place
					
	0.0601467, 0.0174497	0.0474449, 0.00808426	0.0462631, 0.00809828	0.0265373, 0.0105984	0.0460976, 0.00576892
					
	0.0212676, 0.018316	0.0173448, 0.0212341	0.0231289, 0.0146192	0.0219564, 0.0135792	0.0220843, 0.0127368
					
	0.0288468, 0.0273755	0.0283889, 0.015082	0.0233875, 0.0161683	0.0270073, 0.0139235	0.0278969, 0.0128096
					
	0.15266, 0.0131342	0.0908044, 0.00846512	0.075306, 0.00765539	0.0517053, 0.00898845	0.0620719, 0.00718558

7 Conclusions

In this paper a hybrid ensemble of diverse classifiers is proposed and analyzed for the logo and trademark recognition problem. Four member classifiers are used in the ensemble. These are: the structural tensor phase orientation histograms, computed in the morphological scale-space, the novel Hausdorff distance joined with the log-polar correspondence of the image patches around corner points, the affine moment invariant base classifier, as well as the color discriminating classifier. As shown in experiments, thanks to the diversity of the used member classifiers, the ensemble achieves much higher accuracies, not attainable for any of the member classifier when operating alone. Also important is a flexible architecture of the proposed system which allows composition of the classification modules in the serial response narrowing

fashion, or in a parallel operation, depending on the recognition tasks. In the first stage, the system was tested mostly to provide information on its performance with the geometrically transformed and distorted patterns. In this respect the system shows high robustness and compares favorably to other reported systems in this domain. In the second step, the generalization properties of the system were assessed as well. These are important to facilitate the task of searching similar logo or trademark symbols by users. In this aspect, the system outputs a fuzzy similarity measures. Finally, the proposed method is robust in terms of response time and occupied memory.

Acknowledgments This work was financially supported by the Polish National Science Centre NCN, contract no. DEC-2011/01/B/ST6/01994.

References

- Aherne FJ, Thacker NA, Rockett PI (1998) The Bhattacharyya metric as an absolute similarity measure for frequency coded data. *Kybernetika* 34(4):363–368
- Bhattacharyya A (1943) On a measure of divergence between two statistical populations defined by their probability. *Bull Calcutta Math Soc* 35:99–110
- Ballan L, Bertini M, Del Bimbo A, Jain A (2008) Automatic trademark detection and recognition in sport videos. *ICME 2008*
- Bay H, Ess A, Tuytelaars T, Van Gool L (2008) SURF: speeded up robust features. *Comput Vis Image Underst* 110(3):346–359
- Bigun J (2006) *Vision with direction. A systematic introduction to image processing and computer vision*. Springer, Berlin
- Brox T, Boomgaard van den R, Lauze F, Weijer van de J, Weickert J, Mrázek P, Kornprobst P (2006) Adaptive structure tensors and their applications. In: Weickert J, Hagen H (eds) *Visualization and processing of tensor fields*. Springer, Berlin, pp 17–47
- Bugaj M, Cyganek B (2012) GPU based computation of the structural tensor for real-time figure detection. In: *Proceedings of the 20th international conference on computer graphics, visualization and computer vision (WSCG'12)*, Czech Republic
- Chen J, Leung MK, Gao Y (2003) Noisy logo recognition using line segment Hausdorff distance. *Pattern Recognit* 36:943–955
- Cyganek B (2013) *Object detection and recognition in digital images. Theory and practice*. Wiley, New York
- Cyganek B, Siebert JP (2009) *An introduction to 3D computer vision techniques and algorithms*. Wiley, New York
- Cyganek B, Wiatr K (2011) Image contents annotations with the ensemble of one-class support vector machines. In: *International conference on neural computation theory and applications*, 24–26 October, Paris, France
- Dai S, Huang H, Gao Z (2009) Vehicle-logo recognition method based on Tehebicchef moment invariants and SVM. In: *Software Engineering, WCSE'09*, pp 18–21
- Diligenti M, Gori M, Maggini M, Martinelli E (2001) Adaptive graphical pattern recognition for the classification of company logos. *Pattern Recognit* 34:2049–2061
- Doermann DS, Rivlin E, Weiss I (1996) Applying algebraic and differential invariants for logo recognition. *Mach Vis Appl* 9(2):73–86
- Dubuisson M-P, Jain AK (1994) A modified Hausdorff distance for object matching. In: *Proceedings of the 12th IAPR international conference on pattern recognition*, vol 1, pp 566–568
- Flusser J, Suk T (1994) Affine moments invariants: a new tool for character recognition. *Pattern Recognit Lett* 15:433–436
- Flusser J, Suk T, Zitová B (2009) *Moments and moment invariants in pattern recognition*. Wiley, New York
- Gori M, Maggini M, Marinai S, Sheng JQ, Soda G (2003) Edge-backpropagation for noisy logo recognition. *Pattern Recognit* 36:103–110
- Huttenlocher DP, Klanderma GA, Rucklidge WJ (1993) Comparing images using the Hausdorff distance. *IEEE Trans Pattern Anal Mach Intell* 15(9):850–863
- Jackway PT, Deriche M (1996) Scale-space properties of the mutliscale morphological dilation-erosion. *IEEE PAMI* 18(1):38–51
- Jain R, Doermann D (2012) Logo retrieval in document images. In: *10th IAPR international workshop on document analysis systems*
- Krawczyk B, Woźniak M, Cyganek B (2014) Clustering-based ensembles for one-class classification. *Inf Sci* 264:182–195
- Kuncheva LI (2004) *Combining pattern classifiers. Methods and algorithms*. Wiley-Blackwell, New York
- Lughofer E, Buchtala O (2013) Reliable all-pairs evolving fuzzy classifiers. *IEEE Trans Fuzzy Syst* 2(4):625–641
- Mao S, Ye M, Li X, Pang F, Zhou J (2013) Rapid vehicle logo region detection based on information theory. *Comput Electr Eng* 39:863–872
- Petrovic VS, Cootes TF (2004) Analysis of features for rigid structure vehicle type recognition. In: *Proceedings of the BMVC*
- Pham TD (2003) Unconstrained logo detection in document images. *Pattern Recognit* 36:3023–3025
- Phan R, Androutsos D (2010) Content-based retrieval of logo and trademarks in unconstrained color image databases using color edge gradient co-occurrence histograms. *Comput Vis Image Underst* 114:66–84
- Polikar R (2006) Ensemble based systems in decision making. *IEEE Circuits Syst Mag*, pp 21–45
- Psyllos AP, Anagnostopoulos C-NE, Kayafas E (2010) Vehicle logo recognition using a sift-based enhanced matching scheme. *IEEE ITS* 11(2):322–328
- Yang C-HT, Lai S-H, Chang L-W (2007) Hybrid image matching combining Hausdorff distance with normalized gradient matching. *Pattern Recognit* 40:1173–1181
- University of Maryland, Laboratory for Language and Media Processing (LAMP), Logo Dataset. <http://lamp.cfar.umd.edu>, (downloaded 2013)
- Viola P, Jones M (2004) Robust real-time face detection. *Int J Comput Vis* 57(2):137–154
- Woźniak M (2013) *Hybrid classifiers: methods of data, knowledge, and classifier combination. Studies in computational intelligence*. Springer, Berlin
- Zhu G, Doermann D (2007) Automatic document logo detection. In: *The 9th international conference on document analysis and recognition (ICDAR 2007)*, pp 864–868, Curitiba, Brazil
- Zokai S, Wolberg G (2005) Image registration using log-polar mappings for recovery of large-scale similarity and projective transformations. *IEEE Trans Image Process* 14(10):1422–1434