Partial Matching Using Set Exclusion Criteria:
Applied to livestock brand retrieval

Waldemar Villamayor-Venialbo, Horacio Legal-Ayala
Facultad Politécnica
Universidad Nacional de Asunción
Campus Universitario, San Lorenzo, Paraguay
e-mail: wvenialbo, hlegal@pol.una.py

Edson Justino, Jacques Facon
Programa de Pós-Graduação em Informática
Pontifícia Universidade Católica do Paraná
Rua Imaculada Conceição 1155, 80215-901, Curitiba, PR, Brasil
e-mail: edson.justino@pucpr.br, facon@ppgia.pucpr.br

***Abstract*—This article presents a partial matching approach based on set exclusion criteria applied to livestock brand retrieval. A set exclusion measure function, depending on local shape features, is formulated. This exclusion measure helps to determine if a registered brand is a subset of another existing brand, or *vice-versa*. Local features are obtained taking account the relative spatial distribution of brand components. Experimental results have shown the accuracy of the methodology in cattle brand identification.**

***Keywords*-Partial matching; set exclusion criteria; exclusion measure function; livestock brands; handwritten symbols; invariant features; pattern recognition; image retrieval.**

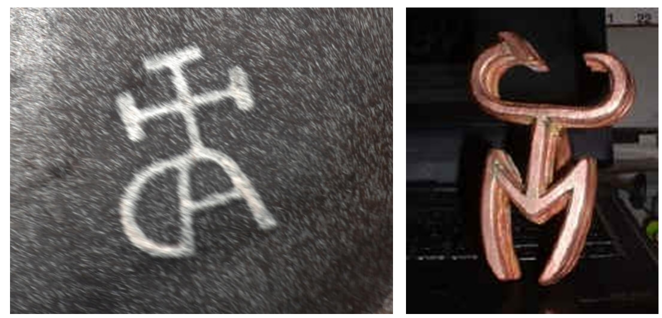
I. Introduction

Visual information retrieval has become an active research area in recent years due to the increasing needs in many application fields. The management of cattle brand databases maintained by government livestock offices represents one of these areas. The correct registration of cattle brands is a major issue in countries with an old ranching tradition due to large brand image databases. Brand inspector officers must prevent frauds ensuring that a particular brand cannot be converted in another brand by adding some extra components to its design [1].

New methods such as implanted RFID chips are positive identifiers, but stamped brands are a permanent and definitive mark of ownership, they are highly visible and hard to alter; actually, branding continues being the *de facto* method to prove ownership of lost or stolen animals in traditional cattle raising countries [2].

Some government livestock offices establish brand reading nomenclature methods that allow database searches by string codes [3], [4], [5]. Others establish their branding rules in a way that brand image database records can be easily retrieved by standard optical character recognition (OCR) techniques [6].

However, we center our interest on the odd hieroglyphic cattle and horse brands used in regions influenced by the Spanish culture. These brands are stamped on animals using hot or freeze pre-shaped branding irons. This type of cryptic brands is composed by strokes of uniform thicknesses. These strokes form letters, numbers, and symbols of any kind; very often they appear overlapped, Fig. 1.



A. B.

Figure 1. Livestock branding: A brand stamped on the animal (A) and a branding iron (B).

There is no published scientific literature on this subject, except for references [7], [8], [9] to some currently working system based on methods discussed above with no further details.

Standard approaches for logo and trademark recognition [10], [11] are not suitable for this kind of brands; by the nature of brands and the problem itself. The same could be said about handwriting recognition techniques; moreover they are more permissive that one can wish with respect to variations in design [12], [13].

Problem Formulation

The challenge consists in identifying candidate brands that may include or be included in existing brand designs, *i.e.*, retrieve from the database any registered brand that could be converted in the candidate brand, or *vice-versa*. The method should be efficient to manage large number of livestock brands. To understand how we face the problem, one case of brand registration procedure will be explained.

Regulations imposed by government livestock offices establish a minimum and maximum size for a brand [14]. When registering a new brand, the applicant is required to draw by hand the desired design for his brand. The sketch is drawn inside a bounding box having the proportions determined by law. While small changes in orientation are permitted, the rules define limits to the acceptance of rotated designs. *E.g.*, two brand designs in which one is rotated $45°$ or more will be considered distinct brands. The same should be argued about reflected, slanted, and shrunk or stretched designs. On the other hand, global changes in scale are not considered distinctive, provided that the sketch is intelligible.

Eventually the sketch of an accepted brand is digitized by the officer and stored in the brand database. In order to systematize the acceptance process the following scenario arises: we should be able to retrieve from the database any brand that could be considered a subset or superset of the candidate brand. At this point we confront two inconveniences: 1) the system must ignore any combinations of the changes cited above, and 2) there must be a restriction in the size of the allowable changes. In this work we will present a solution to the problem described in 1 above.

The article is organized as follows. Section II describes the background of our partial matching approach. Section III provides details on the implementation of the proposed model. Experimental results using real and synthesized cattle brand images are discussed in Sect. IV. Section V gives the conclusion and a few words about future work. Finally, Sect. VI is an Appendix showing the results of intermediate processes, working with real examples of cattle brand samples.

II. Partial Matching

A. Working with Local Shape Descriptors

In our particular case, cattle brand registration is done by handwriting the brand on a graphics table within a bounding box. It means that two similar brands may be drawn with slight differences in size, proportion, orientation, and slanting. In this work, these differences are approximated by an unknown affine transformation.

Due to the nature of the digitalization equipment, the problem is studied in the context of low-noise gray-level images; no other noise is added to the source image.

The proposed partial matching scheme is based on a set exclusion measure function. The formulation of such measure function relies on the comparison of descriptors, obtained by decomposing the brand in some characterizing components.

Each component forming the brand is analyzed separately to extract a set of local descriptors. As we anticipate, we consider these descriptors invariant under affine transformations, *i.e.*, two components related by an affine transformation will be mapped to the same point in our descriptor space.

B. Choosing the Right Components

Strokes as brand components: To perform the decomposition, a natural choice is to consider the primitive strokes forming the cattle brand as the working elements. We define a primitive stroke as a continuous, open or closed, curve or polygon. From now on, every time we refer to a stroke, we are referring to a primitive stroke.

However, matching livestock brands by only taking account each individual stroke, separately, is not enough to distinguish two brands with similar design.

For instance, if we consider their strokes separately, the brand to the left in Fig. 2A could be considered a superset of the brand to the right, the shaded stroke is present only in the brand to the left. Strokes f_1, f_2, g_1 and g_2, compared in 2B-C, are exactly the same, in each case. In 2D, f_3 is the reflection of g_3; reflection is an affine transformation, hence, our descriptor space cannot distinguish between them.

Relative spatial distribution of strokes: The adopted solution consists in using pairs of strokes as brand components. Descriptor values computed from the union of two strokes, preserving their position relative to each other, will be enough to distinguish both brands in Fig. 2A.

Figures 3B–D illustrates the only three matching pairs of strokes to carry out the comparison. Among the three pairs, only two of them are similar 3B and 3C. In 3C, again, one pair is the reflection of the other one. In 3D, we see that f_3 and g_3 appears reflected in each brand, but f_1 and g_1 does

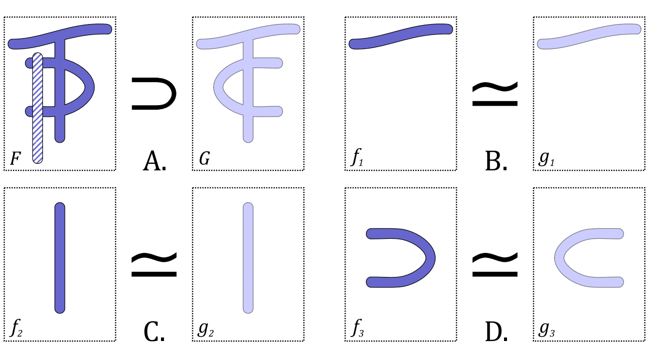


Figure 2. Matching brand elements considering individual strokes.

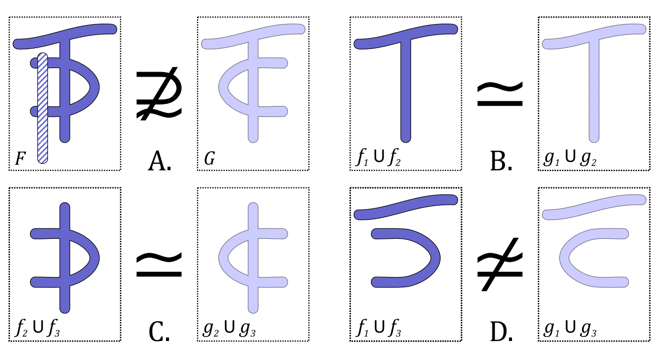


Figure 3. Matching brand components considering pairs of strokes.

not. This geometric fact should be enough to label these brands as different.

C. Brand Classification

1) Set of descriptors of a brand: Considering a brand $F \subset \mathbb{R}^2$ as a countable set of strokes, let $f_i \in F$ represents the i-th stroke in that brand. The set of descriptors of F is defined as

$$S_F := \bigcup_{i=1}^{N-1} \bigcup_{j=i+1}^{N} \mathcal{I}(f_i \cup f_j), \quad N > 1, \tag{1}$$

where $\mathcal{I}\colon F \to \mathbb{R}^n$ is some shape feature extractor operator and N is the number of strokes in F. We do not consider here the trivial case when $N = 1$, *i.e.*, a brand composed uniquely of one isolated stroke.

2) Best association between sets of descriptors: Now, let S_F and S_G represent the descriptor sets of two brands F and G, respectively. Note that $S_F \supsetneq S_G$ implies $F \supsetneq G$, *i.e.*, brand G could be converted in F by adding the strokes in $F \setminus G$.

Consider the countable collection $\mathcal{P}_{FG}$ of injective (one-to-one) correspondences with the form

$$P_{FG} = \{\, (x, y) \mid x \in S_F,\ y \in S_G \,\}. \tag{2}$$

In the case of two similar brands, only one of the correspondences $P_{FG} \in \mathcal{P}_{FG}$ will depict the correct association of sets S_F and S_G.

In order to determine if S_F is a subset or superset of S_G, we must find that correspondence. To solve this issue a preselection on the sets $P_{FG} \in \mathcal{P}_{FG}$ using a minimum distance classification criterion will be performed.

Consider the linear combination of the distances between each pair of descriptors $(x, y) \in P_{FG}$ given by

$$d(P_{FG}) := \sum_{(x,y)\in P_{FG}} \|x - y\|, \tag{3}$$

where $\|\cdot\|$ is some norm in the descriptor space $\mathbb{R}^n$, *e.g.*, the L_2 norm. Note that $Q_{FG} \in \mathcal{P}_{FG}$ contains the best pairing of members in S_F and S_G if, and only if it fulfills the best association condition

$$d(Q_{FG}) = \min_{P\in\mathcal{P}_{FG}} d(P). \tag{4}$$

Once Q_{FG} have been determined, we must verify that the association it describes is good enough to consider if S_F, or S_G, is included in the other.

3) <u>Set exclusion criteria</u>: One way to determine that a brand does not contain, nor it is contained in another brand, is establishing an exclusion measure $M\colon \mathbb{R}^2 \times \mathbb{R}^2 \to [0,\ 1]$. Such measure between two brands must yields 0 whenever a brand is a subset of the other, and 1 when the brands are disjoint sets. Values between 0 and 1 will indicate that brands are not disjoint sets but neither one is subset of the other one; *i.e.*, their intersection is not an empty set. We will now formulate one of such measures which will be the workhorse of our partial matching approach.

Let m and n be the number of elements in S_F and S_G, respectively, we define our set exclusion measure by

$$M(F, G) := 1 - \frac{1}{r}\sum_{i=1}^{r} \chi_i\,, \tag{5}$$

where $r = \min(m, n)$, and the characteristic operator χ_i is defined as

$$\chi_i := \begin{cases} 1, & \text{if } \|x_i - y_i\| \le \theta\,, \\ 0, & \text{otherwise}\,, \end{cases} \qquad (x_i, y_i) \in Q_{FG}\,. \tag{6}$$

As in (3), $\|\cdot\|$ is some underlying norm from the descriptor space, and $\theta \ge 0$ is a threshold parameter determining the maximum allowed distance to consider similar two corresponding components from F and G.

Measuring exclusion: The set exclusion measuring function (5) can yield any value from Table I.

$M(F, G) = 0$ means that F or G, or both, contains all the strokes of the other; it will be $F = G$ only if $m = n$.

When $0 < M(F, G) < 1$, F and G have some strokes in common, but also both have strokes that does not exists in the other.

If $M(F, G) = 1$, F and G have no strokes in common. *I.e.*, both have completely distinct designs.

Summarizing, if $M(F, G) \neq 0$, neither G nor F could be converted in the other.

D. Registering Brands

Regarding the values given by the set exclusion measure (5) summarized in Table I, and considering G as any brand from the database of registered brands, a registration system must reject any candidate brand F for which $M(F, G) = 0$.

Figure 4 illustrates the only four scenarios a registration system could face, and the values given by the set exclusion measure for each case. The lower row shows the candidate brands, labeled $F_{1,\ \ldots,\ 4}$. A brand from the database, labeled G, is in the mid-upper part of the figure. From left to right: F_1 is, clearly, a superset of G, the shaded stroke is not present in G; F_2 is a subset of G, since it lacks one stroke from G; F_3 has three strokes in common with G, but the one it lacks from G and the shaded stroke in F_3 makes it impossible the conversion between them; finally, F_4 have a completely distinct design from G.

According to the exclusion criterion both brands, F_1 and F_2, must be rejected since $M(F_1, G) = M(F_2, G) = 0$.

Table I
CONTINGENCY OF THE SET EXCLUSION MEASURE

Relation	Exclusion value
$S_F \supseteq S_G$ or $S_G \supseteq S_F$ $\implies$	$M(F,G) = 0$
$S_F \not\supseteq S_G$ and $S_G \not\supseteq S_F$ $\implies$	$0 < M(F,G) \le 1$

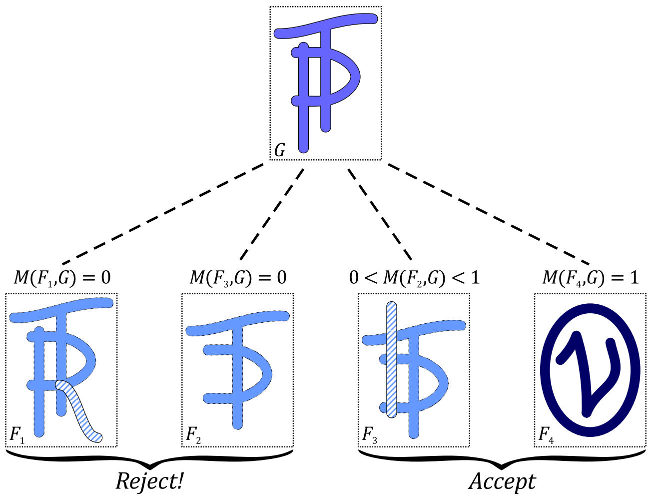


Figure 4. Values of the set exclusion measure for various scenarios.

At the same time, F_3 and F_4 may be accepted, because $M(F_3, G) \neq 0$ and $M(F_4, G) \neq 0$.

III. Implementation Discussion

We will not give a detailed description of the auxiliary stages of our implementation. What follows is just an overview—properly referenced—explaining what is done before the matching stage is performed. Sect. VI illustrates some example cases with real brand images, where the output of these preliminary processes and its intermediate stages are shown.

A. Image Preparation and Brand Segmentation

The experiments have been applied onto 8-bit gray-level images. Brand sketches digitized with the graphics table are noise free. No noise has been added, neither in the images background nor in the brands strokes.

First, query and database images were binarized using Otsu's threshold algorithm [15]. Then, each pixel in the strokes are analyzed and labeled with its degree of freedom using techniques described in [16]. The original strokes are broken in their junction regions and each primary stroke is isolated and labeled. The *orientation space projection* technique described in [17], [18], was chosen to reconstruct all continuous strokes, separating overlapping strokes. This technique can easily handle X- and T-type strokes junctions. L- and V-type strokes are also broken in their vertices, these segments are recombined because we consider them single strokes.

B. Feature Extraction

For the feature extraction stage we have several approaches to choose from. Standard *affine moment invariants* (AMI) described in [19] and generalized in [20]. Two methods based on a new image transform called *multiscale autoconvolution* (MSA) [21] and *spatial multiscale affine invariants* (SMA) [22]. And a novel approach based on the combination of the theory of the AMI, MSA and SMA methods, the *generalized affine moment invariants* (GAM) [23]. All descriptors extracted by those methods are invariant under affine transformations.

Initially we made some testing with AMI features, then we switched to the SMA method for its robustness and its low complexity at the implementation time—with respect to MSA and GAM models—and because AMI possess considerable weaknesses [22] confirmed by our experiments.

C. Partial Matching Classifier

Although preliminary experiments was done using the L_∞ and L_1 norms, the performance tests was carried out using the Euclidean L_2 norm—which bring better results—in $\mathbb{R}^{35}$, the descriptor space supplied by the SMA operator [22].

A graph path minimization method is used to determine the components correspondence fulfilling the best association condition (4). The implementation of the set exclusion measure function (5) and the characteristic operator (6)—the workhorses of our partial matching scheme—is straightforward.

The value for the parameter θ in (6) was determined experimentally to achieve a good trade-off between sensitivity and selectivity. The performance was checked out using values for θ in the range $[0.004,\ 0.1]$. Decreasing the value of θ increase the classifier's selectivity, and increasing θ increase its sensitivity.

IV. Experimental Results

A. Database of Real Brands

To quantitatively evaluate the accuracy and efficiency of the proposed partial matching scheme, we perform brand retrieval according to exclusion criteria summarized in Table (I). A set of 315 randomly selected real cattle brand images was established to setup a database and carry out the testing. This database was preprocessed to populate the knowledge base with the sets of descriptors. Figure 5, shows a slice sample of the brands in our database.

B. Synthesized and Real Query Brands

Fourteen test sets containing synthesized query brand images have been prepared. The test sets were grouped into three categories:

1) Five sets for trivial tests totalizing 1563 query images containing: verbatim copies of the brands in the reference database, and the same brands with add-on aggregates consisting of external arcs and bars, circles and rectangles bounded to the brands.
2) Six sets with 1650 query images synthesized from the original brands by applying affine transformations. They include one set combining transformed brands with add-on aggregates.

Figure 5. Real brand samples from our cattle brand database.

3) Three sets totalizing 749 query images including: database brands with add-in aggregates (extra strokes overlapping the original brand design), partial brands constructed by removing out one stroke from the database brands, a rejection set containing brands with no counterpart in the database; 35 real brand images were also included in this last set.

In order to isolate the inherent error in the retrieval model from external influences, those images for which the segmentation procedure did fail were discarded, even though, the amount of error is reported when it is applicable. The correctness rate of the segmentation process varies between 82% and 86%, consistently with those values given in [16], [17].

C. Evaluation Criteria

The metrics and criteria given by Eq. (7) and Table II, respectively, were adopted to evaluate the intraclass and overall retrieval performance [24]:

$$P = \frac{tp}{tp + fp}\,, \qquad R = \frac{tp}{tp + fn}\,. \tag{7}$$

Table II
CONTINGENCY FOR THE RETRIEVAL TEST SETS

	$F \supseteq G$ or $G \supseteq F$	$F \not\supseteq G$ and $G \not\supseteq F$
$M(F,G) = 0$	True positives	False positives
$M(F,G) \neq 0$	False negatives	True negatives

Table III
BEST INTRACLASS RETRIEVAL PERFORMANCE FOR THE PROPOSED MODEL

	Test set	P	R	S.E.
A.	Verbatim	1.000	1.000	0.0%
	Add-on	1.000	1.000	0.0%
B.	Displaced	0.971	0.971	13.3%
	Rotated	0.977	0.970	18.1%
	Scaled	0.946	0.902	13.7%
	Stretched	0.988	0.952	14.9%
	Sheared	0.992	0.992	16.2%
	Combined	0.973	0.954	16.5%
C.	Add-in	0.980	0.918	3.7%
	Partial	1.000	1.000	12.4%
	Rejection	1.000	1.000	N.A.

P measures the *precision* of the proposed model, an indicator of its selectivity, R measures the *recall* value, an indicator of the sensitivity of the model, tp is the amount of query brands correctly classified, *true positives*, fp is the amount of misclassified query brands, *false positives*, and fn the amount of not retrieved query brands, *false negatives*.

D. Performance of Partial Matching

Table III summarizes the results of feeding the prototype program with the test sets. The column labeled "S.E." list the amount of segmentation error, the P and R values are not affected by this error.

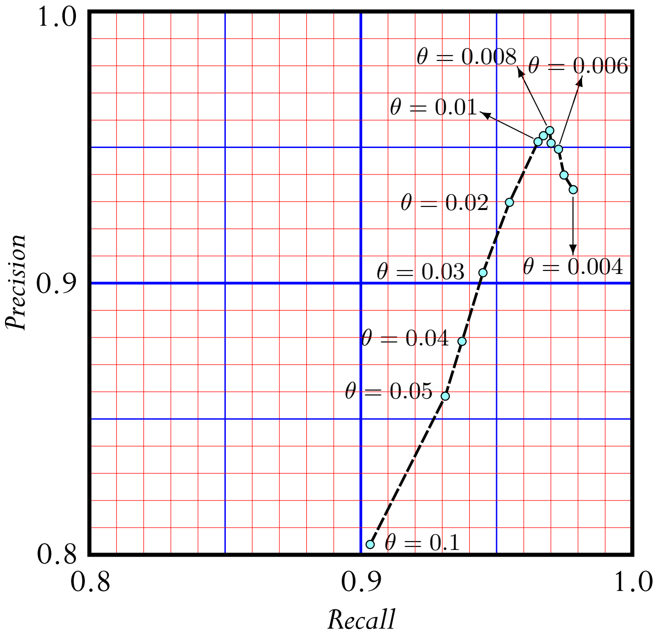


Figure 6. Overall retrieval performance for the proposed model. The leftmost measure is for $\theta = 0.1$, and the rightmost for $\theta = 0.004$.

The trivial tests (group A) achieved $P = R = 1.000$ for all values of θ in our test interval $[0.004,\ 0.1]$. For the sets with brands undergoing affine transformations (group B), the best performance was obtained for $0.006 \leq \theta \leq 0.01$. Brands with add-in strokes (group C) yield $P \geq 0.950$ for $\theta \leq 0.006$ and $R \geq 0.900$ for $\theta \leq 0.005$. On the other hand, partial and rejection brands (group C) attained good precision for values of $\theta \leq 0.01$, and $R = 1.000$ always.

The low performance achieved with the scaled version is because the brand images are not scaled in the true sense but resized versions of the original images preserving the original width of the strokes. For the stretched and sheared variations we attained better results, since in those cases the scaling only affects one axis of the image. The larger segmentation error occurs with rotated brands, the source of the error is, indeed, the sensitivity of the segmentation technique to the orientation of the strokes.

The overall retrieval performance is shown in Fig. 6, presented as the *recall-precision* graph [24]. This includes the results of all test sets with the exception of the trivial cases. It can be seen that good performance (high rates of recall and precision) is accomplished for $0.004 \geq \theta \geq 0.01$.

V. Conclusion and Future Work

In this article, we proposed a partial matching algorithm based on set exclusion criteria. A strategy based on local descriptors extracted from pairs of strokes was used to formulate a set exclusion measure function.

Experimental results, summarized in Table III and Fig. 6, have shown the efficiency of the partial matching methodology in livestock (cattle) brand retrieval. We believe its functionality could be applied to other image analysis and pattern recognition areas, offering new possibilities for similar applications.

For future work, we aim to enhance the overall accuracy of the system improving the stroke segmentation stage with new approaches, like the method presented in [25], [26], and implementing more powerful feature extraction techniques such as those in [21], [23].

VI. Appendix

A. Real World Examples

This section will show working examples with real cattle brands. Fig. 7 illustrates the intermediate steps of the segmentation stage. The upper row shows a real brand from the cattle brand database, the “JL Ranch” brand. In the lower row we modified the JL brand, the original marks were slightly slanted and three extra strokes have been added; this is a fake “hat HF combined” brand.

In 7A the original digitized JL and the modified HF brands are shown; Fig. 7B shows the result of binarization using the Otsu’s thresholding algorithm; Fig. 7C are the result of labeling each pixel in the stroke with its degree of freedom, this identifies terminal (darker), regular and junction (lighter) regions; and Fig. 7D shows the labeled primary strokes, after the original strokes were broken at their junction regions.

Next, each junction region and its surrounding are mapped into a 3D orientation space (OS). All primary strokes which results connected in the OS must be reconnected after projecting back the OS mapping onto the image plane. The resulting sets of pixels are considered the continuous primitive strokes. Figures 7E–I illustrates the results of these processes. Some irregularities in the stroke contour appear

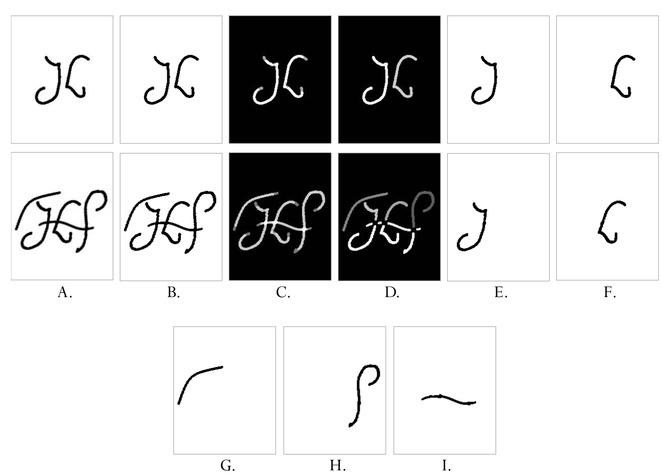


Figure 7. Brand segmentation: Identification and isolation of primary strokes (A–D), reconstruction and extraction of continuous primitive strokes (E–I).

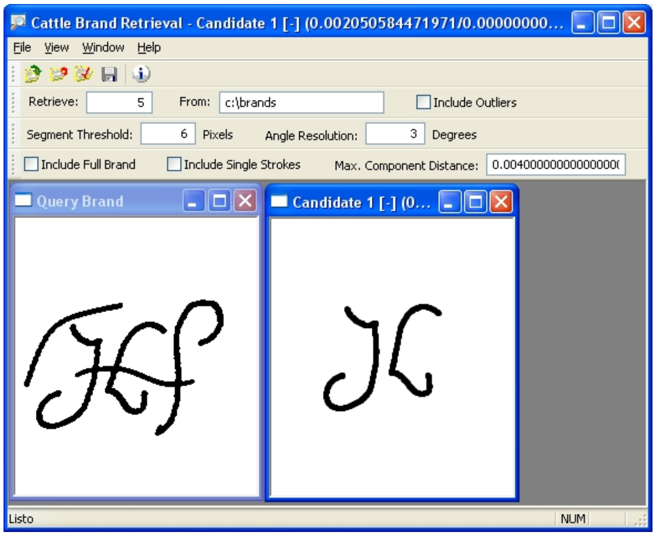


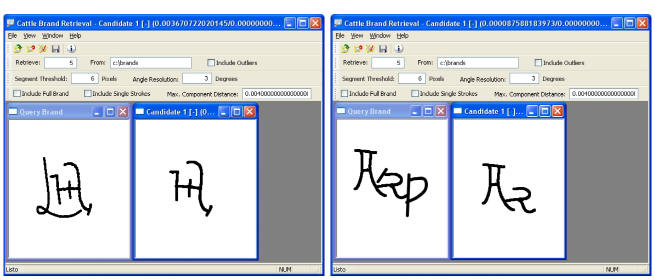
Figure 8. Example case: The exclusion measure for the query F and database brand G yields $M(F, G) = 0.00$, the value for the distance between the only matching pair components is $\|x_i - y_i\| = 2.05 \times 10^{-3}$.

where there were the junction regions. This technique is unable to handle tangent curves [17].

Figure 8 shows how the correct brand is identified for this example case. The response time varies with the complexity of the query brand, but it normally takes less than two seconds to a 3.0 GHz Pentium IV to find a matching candidate with our small database; in this situation most of the time is consumed by the segmentation procedure.

The target database is supposed to have about of 80 000 registered brands, in such a case the time consumed by the segmentation procedure will be relegated to a second plane.

Figure 9 illustrates the output of our implementation for two other examples. For these and the previous example we set $\theta = 4.0 \times 10^{-3}$.



A. B.

Figure 9. More examples: A) In this case the original brand was slightly rotated and shrunk horizontally, and then a new stroke was added. The exclusion measure is $M(F, G) = 0.00$, and the mean value for the distance between matching pair components is $\|x_i - y_i\| = 3.67 \times 10^{-3}$. B) In this example we just add three new strokes to the unmodified original brand. The exclusion measure is $M(F, G) = 0.00$, and the mean value for the distance between matching pair components is $\|x_i - y_i\| = 8.76 \times 10^{-5}$.

B. System Prototype

Figure 10 shows a diagram of the complete system prototype, starting since a new query is digitized all the way to its acceptance or rejection. Whenever a new brand is accepted, its digitized image and its set of descriptors are stored in the brand image database and in the descriptors set knowledge base, respectively, completing the registration process. The prototype was implemented in the C++ language.

REFERENCES

[1] The Cattlemen's Texas Longhorn Registry, "Cattle brand registration procedures," Website of the The Cattlemen's Texas Longhorn Registry, December 2008, http://www.ctlr.org.

[2] Texas and Southwestern Cattle Raisers Association, "Law enforcement and livestock inspection," Website of the Texas and Southwestern Cattle Raisers Association, December 2008, http://www.texascattleraisers.org.

[3] The Cowboy Showcase, "Reading and understanding livestock brands," The Cowboy Showcase Website, December 2008, http://www.cowboyshowcase.com/brands.htm.

[4] The Government of Saskatchewan, "Brand registration procedures and specifications," Website of the Department of Agriculture, Dec. 2008, http://www.agr.gov.sk.ca.

[5] The State of New Mexico, "Branding code and regulation," Website of the New Mexico Livestock Board, December 2008, http://www.nmlbonline.com.

[6] L. A. Martínez, "Livestock brand recognition system," World Intellectual Property Organization, Moriarty, NM, USA, Patent International Pub. No. WO 2005/070114 A3, 2005.

[7] The Government of Queensland, "Branding and earmarking livestock in queensland," Website of the Department of Primary Industries and Fisheries, December 2008, http://www.dpi.qld.gov.au.

[8] The State of Nebraska, "Nebraska brand book," Official Nebraska Government Website, December 2008, http://nbc.nol.org/brandbook.html.

[9] The State of Oregon, "Animal health and identification division," Website of the Oregon Department of Agriculture, December 2008, http://oda.state.or.us/dbs/brands.

[10] A. D. Bagdanov, L. Ballan, M. Bertini, and A. del Bimbo, "Trademark matching and retrieval in sports video databases," in *Proceedings of Nth MIR*, 2007.

[11] R. H. van Leuken and J. Austin, "Layout indexing of trademark images," in *Proceedings of the 6th International Conference on Image and Video Retrieval*, 2007.

[12] M. Morita, , R. Sabourin, F. Bortolozzi, and C. Y. Suen, "Segmentation and recognition of handwritten dates," in *Proceedings of the 8th IWFHR*, 2002, pp. 105–110.

[13] S. Ouchtati, B. Mouldi, and A. Lachouri, "Segmentation and recognition of handwritten numeric chains," *Intl. Journal of Information Technology*, vol. 4, no. 1, pp. 37–44, 2007.

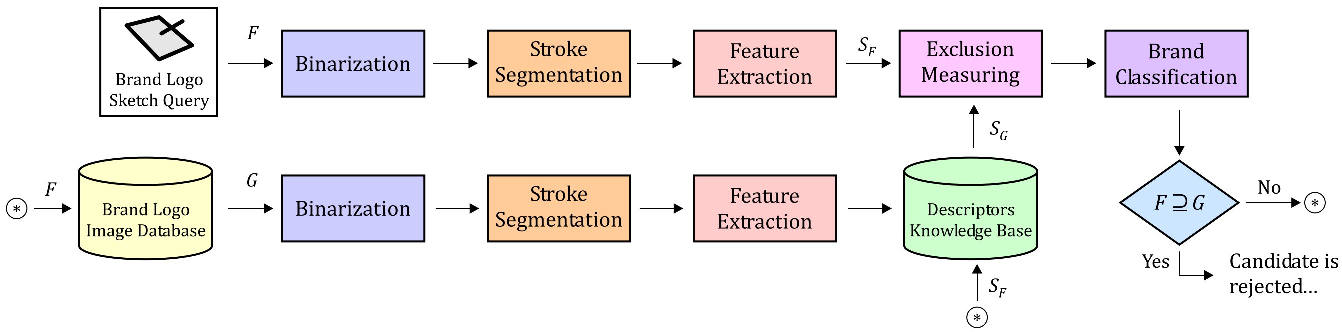


Figure 10. A scheme of the livestock brand registration system.

[14] The Government of Paraguay, "Ley N^{o} 2576: de reinscripción de marcas y señales para todos los propietarios de ganado mayor y menor," Website of the Presidence of the Republic, 2005, http://www.presidencia.gov.py.

[15] N. Otsu, "A threshold selection method from gray-level histogram," *Transactions on System, Man, and Cybernetics*, vol. 8, pp. 62–66, 1978.

[16] R. Cao and C. L. Tan, "Line primitive extraction by interpretation of line continuation," in *Proceedings of the 6th International Conference on Control, Automation, Robotics and Vision*, 2000.

[17] ——, "A model of stroke extraction from chinese character images," in *Proceedings of the 12th International Conference on Pattern Recognition (ICPR)*, vol. 4. IEEE, 2000, pp. 368–371.

[18] J. Chen, Y. Sato, and S. Tamura, "Orientation space filtering for multiple orientation line segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 22, no. 5, pp. 311–317, May 2000.

[19] J. Flusser and T. Suk, "Pattern recognition by affine moment invariants," *Pattern Recognition*, vol. 26, no. 1, pp. 167–174, May 1993.

[20] T. Suk and J. Flusser, "Graph method for generating affine moment invariants," in *Proceedings of the 16th International Conference on Pattern Recognition (ICPR)*. IEEE, 2004, pp. 192–195.

[21] E. Rahtu, M. Salo, and J. Heikkilä, "Affine invariant pattern recognition using multi-scale autoconvolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 27, no. 6, pp. 908–918, 2005.

[22] ——, "A new efficient method for producing global affine invariants," in *Proceedings of the International Conference on Image Analysis and Processing (ICIAP)*. Springer, 2005, pp. 407–414.

[23] E. Rahtu, M. Salo, J. Heikkilä, and J. Flusser, "Generalized affine moment invariants for object recognition," in *Proceedings of the 18th International Conference on Pattern Recognition (ICPR)*, vol. 2: 4. IEEE, 2006.

[24] C. D. Manning, P. Raghavan, and H. Schütze, "Evaluation in information retrieval: Evaluation of unranked retrieval sets," in *Introduction to Information Retrieval*. Cambridge University Press, 2008, ch. 8, pp. 142–145.

[25] W. Villamayor-Venialbo and H. Legal-Ayala, "Stroke segmentation from livestock brand images," in *Proceedings of the 35th Conferencia Latinoamericana de Informática (CLEI)*, 2009.

[26] B. Whited, J. Rossignac, G. G. Slabaugh, T. Fang, and G. Unal, "Pearling: Stroke segmentation with crusted pearl strings," *Pattern Recognition and Image Analysis*, vol. 19, no. 2, pp. 277–283, 2009.