2004－DBS－134　(II)　(36)
2004／7／14

An Interactive Annotation Framework for Image Retrieval

Erwan LOISANT[†,††], José MARTINEZ[†], Hiroshi ISHIKAWA[††], Manabu OHTA[††], and Kaoru KATAYAMA[††]

† University of Nantes, Polytechnic School
†† Tokyo Metropolitan University, Graduate School of Engineering

Abstract Image retrieval is usually based on content information, external annotation, or a combination of the two. While content information is intrinsic to the image and thus is an objective data, external annotation may imply human and be deeply dependent on the culture and sensibility of the annotator. Consequently, two operators would not produce the same annotation for the same image, and both results may not be satisfying for user. However, an image usually contain important information impossible to retrieve from content, thus relying only on content information does not produce satisfying results. To answer to this issue, we propose an annotation framework featuring an explicit separation of objective and subjective informations, objective informations including content based information and a limited number of annotations, the other annotations being considered as subjective information and collected with the help of the user. We also propose an efficient way to apply these subjective annotations to a navigation structure calculated before-hand from the objective informations. The resulting browsing technique brings the relevance of human annotation without the cost of a before-hand annotation, and without denying the robustness of a before-hand calculated structure.

Key words image retrieval, meta-modelisation, user interface

1. Introduction

When designing an image retrieval system, selecting relevant description data to use is along with the navigation technique to use, a major issue. There are mainly two approaches: prefer annotation on the media, or content based information. The first approach usually relies on human annotators who have to choose appropriate keywords describing the best the image they are viewing. They may be limited to a given dictionary or a thesaurus to reduce the polysemy and homonymy problems. The second one is fully automated and based on metric from the image itself: usually color and shape information are used, as well as general informations on the format.

External annotations can get very good results, and is even needed to discover some links: we can not expect a fully automated process to see a similarity between an swimming fish and a cooked fish, however the link is obvious for a human observer.

However human annotations are usually subjective, and this is an issue in information retrieval. It has been constated that two annotators will produce different annotations on the same content, even the same annotator asked twice to annote data at different time will produce different results. Consequently, there is no guarantee that the before-hand annotation will be relevant for a given user.

Moreover, annotations are made of objective annotations, like the author name and subjective annotations, like the objects represented on the image. For example, figure 1. could be simply indexed as "a deer" by a zoologist and with an appropriate thesaurus telling

図 1 A deer

that a deer is a cervine, being itself an animal, most users would be happy with that. However an other user may want to group images in a set of images he would call "cute animals". However, the name of the photographer will not change, and it will always be relevant since a given photographer usually take the same kind of photographs and conforms to his own style.

Therefore, we propose an hybrid schema based on an *objective core* including objective annotations as well as content-based meta-data, and *subjective schemas* based on the needs of a given user.

Subjective schemas will be created by the user himself during the retrieval process. These subjective schemas are to be applied on a navigation structure previously calculated from the objective core, and it results in a retrieval navigation-based retrieval process. We prefer a navigation process [18] to query [11], [6] or relevance-feedback [20] because of its advantages in term of ease of use and performances.

Using users interaction to improve the base has already been studied in the case of retrieval based on query [23]. Offering this possibility in a navigation-based retrieval process will provide similar relevance with the ease and performances of navigation on a before-hand calculated structure.

2. Objective Schema

The *annotation core* and the *content-based model* form together the *objective schema*; it is indeed an hybrid schema including external annotation that may be automated or not according the data source, and content-based metadata that are necessarily automatically extracted.

The objective schema is computed before-hand, is common to all users and is used as a base for indexing and basic retrieval.

2.1 Annotation core

It is very restricted, to ensure that there may be no differences between the objective schema expected by one user and an other. We use a subset of a schema proposed by the web community, and standardized by the World Wide Web Consortium (w3c): the photo-rdf schema [16] part of the Dublin Core initiative.

The schema used is the following:

- title: a short description of the photo.
- creator: the photographer
- publisher: the person or institution making the photo available, often the same as the creator
- contributor: a person who contributed in some way, like scanning the photo or applying a set of filters
- date: the date and time the photo was taken
- coverage: the location shown on the photo.

Note that several fields are discarded from the photo-rdf proposal; some of them because they are irrelevant to image retrieval, and some of them because they represent subjective annotations that would be different from an annotator to an other.

In the case of the acquisition source provide it, those fields may be filled automatically, for example using the exim information of digital camera. In any other case it has to be manually filled or left blank.

2.2 Content-based model

When selecting content-based information to be used in our schema, we must keep in mind that these informations need to make sense to a human observer [22] or at least can be interpreted by a semantic description.

This models is focused on structural information and physical information: a general segmentation of image and dominant colors on these parts. Some studies also work on *shape* or texture *Liu-Picard-96* information, but for algorithmic complexity reason we prefer not to use shape or texture.

2.2.1 Color representation

Our objective schema is mainly based on color, an information that has been shown to provide good semantics for a low process cost and thus has been largely used in content-based retrieval systems [9] [12] [8].

Color is know to be a tri-dimensional parameter, however several models exist and we need to choose the more appropriate one [13] [15].

For computer image manipulation, technical color models like RGB or CMYK are preferred. Those models reflect the way pixel's colors are produced by the rendering device. Those technical models however are not suitable for human intuitive color representation. The *pink* color for instance is not easy to describe in term of red, green and blue combination. More accurate models, also said *perceptual* models are then used. The first of those models was proposed by A.H. Munsell in 1915.

The HSV color model, used in this work, is recognized to be one of the most perceptually evident for users [13]. HSV stands for Hue, Saturation and Value. All those components are immediately understandable: hue is the color on the rainbow, saturation the strength of the color (being vivid or pale) and value the luminosity, dark or light.

Using the HSV model, we define a segmentation of the color space based on a subdivision on each component. This subdivision is based on Zadeh's fuzzy logic [5], [10].

For instance, a color may be "dark unsaturated blue".

2. 2. 2 General Shape

An other content-based information used is the *general shape*, size and orientation. Just as on color information, fuzzy subsets are defined on the surface of the image and on the measure:

$$Orientation = log\frac{width}{length}$$

Usage of the logarithm is motivated by the necessity to keep the measure linear compared to human perception. Figure 2. 2. 2 shows the subdivision of the orientation measure.

2. 2. 3 Segmentation

An image segmentation is used to allow a more accurate description of image colors. We want to separate semantically different objects appearing on the image. A segmentation algorithm could be used to identify shapes by considering color or textures differences [25], but these algorithms are costly and will not ensure that region computed are actually real-world objects. Most of the time, a real-world object has several parts of different colors and textures and segmentation algorithms are not able to group these parts.

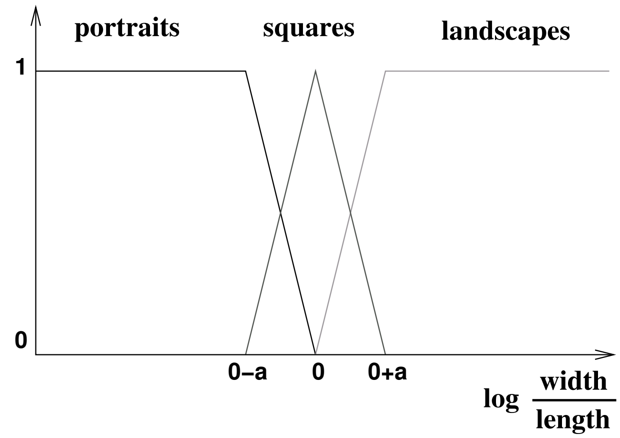


図 2 Fuzzy subsets on orientation

Instead, we consider general rules of photographic pictures. The main subject often stands in the center and the surrounding areas represent the image background. In a landscape picture for instance, the sky is likely to have blue or gray hues, while the ground will probably be green. We use a five zone segmentation. The center zone covers 40% of the total surface and the four surrounding zones are trapezoids whose wideness is 15% of the image wideness.

3. Indexing and Basic Retrieval

To index data using the objective core, most authors use a multi-dimensional indexing [4], [3] [7]. These technics are useful for a query based retrieval, but for navigation we prefer a structure that may be used directly to perform the navigation. This graph structure, the Galois' lattice, has already been used as a navigation structure [19].

3. 1 Galois' lattices

This part gives a quick introduction to Galois' lattices, mainly to precise axioms that we will have to respect while applying filter on it and to introduce notation that will be used later. Interested reader may refer to [19] where Galois' lattices applied to image retrieval are detailed.

A Galois' (or concept) lattice is a mathematical structure that has been largely exploited in the field of knowledge discovery [14] [24]. It can be defined whenever there is a binary relation, in our case between *images* and their associated *meta-data*:

$$R : \mathcal{I} \times \mathcal{D} \tag{1}$$

where $\mathcal{I}$ is the set of images, and $\mathcal{D}$ is a set of descriptions. Note that a Galois' lattice can be defined only over discrete domains.

A lattice being a directed acyclic graph featuring a minimal node (inf) and a maximal node (sup), a Galois' lattice is a special kind of lattice derived from a binary relation.

Each node of this graph groups a set of instances, i.e., an *extension*, and a set of descriptions, i.e., an *intention*. From R, one derives the *Galois' connection* between $\mathcal{I}$ and $\mathcal{D}$, which consists in two dual functions, or point of views on R:

$$\begin{array}{llll} r: & \mathcal{I} & \rightarrow & 2^{\mathcal{D}} \\ & i & \mapsto & \{d \in \mathcal{D} | (i,d) \in R\} \end{array} \tag{2}$$

$$\begin{array}{llll} r': & \mathcal{D} & \rightarrow & 2^{\mathcal{I}} \\ & d & \mapsto & \{i \in \mathcal{I} | (i,d) \in R\} \end{array} \tag{3}$$

Intuitively, r gives the description of each image, i.e., its associated meta-data. In contrast, r' gives images featuring a given property.

The resulting graph is oriented according the following partial order:

$$\begin{array}{llll} S: & (2^{\mathcal{D}} \times 2^{\mathcal{I}})^2 & \rightarrow & \{0,1\} \\ & ((X_1, X'_1), (X_2, X'_2)) & \mapsto & (X_1 \subset X_2) \bigwedge (X'_2 \subset X'_1) \end{array} \tag{4}$$

The inf node and the sup node are also defined according this partial order: the inf node will be the smallest property set associated to the largest image set, and the sup node will be the largest set of property set associated to the smallest image set.

Hence, a *class extension* is defined as:

$$\begin{array}{llll} c: & \mathcal{I} & \rightarrow & 2^{\mathcal{I}} \\ & i & \mapsto & \{i' \in I | r(i') = r(i)\} \end{array} \tag{5}$$

Intuitively, we are interested in the set of images that share *exactly* the same description, and moreover *at least* the same description.

The problem of updating a Galois' lattice is not trivial, since it is necessary to generate not only the new pairs and its connections but usually several other pairs needed to respect the Galois' lattice definition. [14] proposes an incremental algorithm that has an exponential complexity in the worst case. However, in most case we experience a linear complexity for adding one instance.

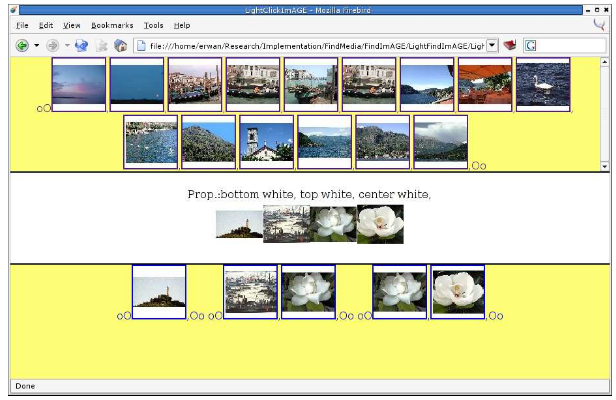


図 3 The retrieval process

A Galois' lattice will be noted

$$G = (\mathcal{N}, \mathcal{E})$$

, where N is a set of nodes and E a set of oriented edges.

3.2 The retrieval process

A direct hypermedia representation [1], [2] of the structure is constructed, and user is to navigate into the graph by clicking on child nodes (to specialize the "query") or on the parent nodes (to generalize).

For our purpose, the advantages of Galois' lattices are numerous.

- First of all, it is very fast to navigate through a graph structure that has been computed off-line. If we neglect the time required to load sample images, navigating from one node to another is optimal, i.e., in $O(1)$. This was one of the main requirements.
- Then, a Galois' lattice is intrinsically a multi-dimensional classification technique. Indeed, no dimension is privileged. Hence, it can be seen as a structure that dichotomizes the hyper-cube associated to the property subsets along any hyper-plane.
- Consequently, the distance from the inf or sup nodes of the graph to any other node is at most logarithmic in the number of used properties.
- Next, this tool is insensitive to correlations. There is no distance computation. If images with a given property (almost) always exhibit another property, then the images will simply be located within the same node.
- Also, this tool helps to correct users' mistakes very easily. Whenever a user selects a direct descen-

dant node, he or she adds implicitly a new constraint. If he or she figures out, much later, when seeing more specific sample images, that this browsing direction is slightly bad, he or she has just to move to a different direct ancestor node. This operation removes a constraint and undoes the erroneous move without having to go back to the point where the "error" actually occurred.

- The Galois' lattice structure easily hides unwanted features. This is a problem that cannot always be taken into account by similarity measures. (A counter-example is Surfimage [21], but the measures are limited to mean and variance of supposed Gaussian distributions.)

There are also some disadvantages.

- The description space associated to a Galois' lattice is exponential in the number of properties. (We easily use several hundreds!) Of course, this should not occur, unless we index such a large number of images. However, if several images share common properties but have unique properties too, then a (localized) exponential explosion appears.
- Also, constructing a Galois' lattice is not an easy task. The time complexity is in $O(n^2)$ where n is the number of nodes. Theoretical improvements on this bound are still unknown, to our knowledge, and algorithmic variants do not achieve actual improvements in the implementations [14].

To answer to these issues, a masking technique [17] has been designed.

4. Subjective Schemas

Subjective schemas are used for advanced retrieval, to allow users to precise the model according their needs. Note that the annotation process is integrated to the retrieval process, and not done before-hand by an annotator.

Figure 4. presents a set of images presenting a semantic link (photographs related to the war in Iraq) but it is impossible for a fully automated process to discover this link. A knowledge of the events in the last years is needed.

There are several motivations to postpone the annotation process to the retrieval time, and not make the annotation before-hand by an expert. First, this process is very costly, as any manual intervention on data.

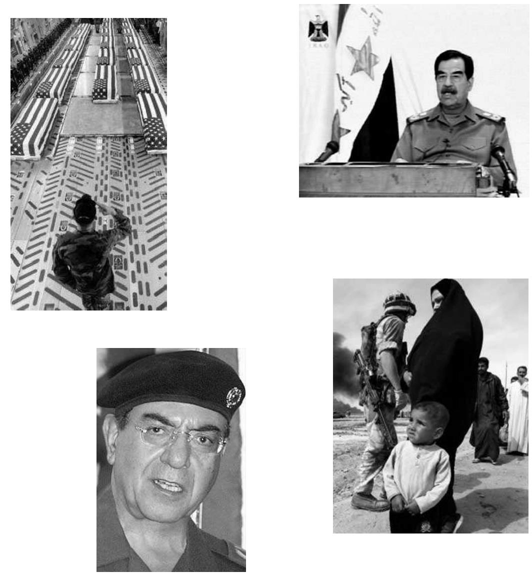


図 4 A set of images visually very different, but that may conceptually be linked by "Iraq"

Secondly, the experts annotators are disconnected from the user needs, and this process is always subjective. It has been constated that two annotators will produce different annotations on the same content, even the same annotator asked twice to annote data at different time will produce different results. Consequently, there is no guarantee that the before-hand annotation will be relevant for a given user.

4.1 Schemas Acquisition

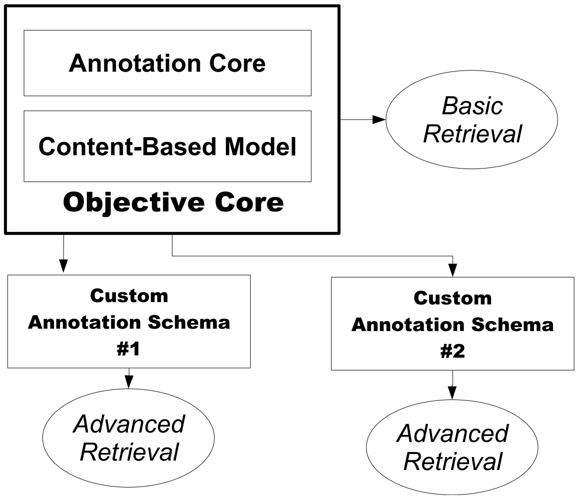
The annotation process is the following: performing a basic retrieval as described in section 3., the user is given by the interface the opportunity to add his own keywords to images. By remaining to the user the keywords he already used, we avoid usage of different words for the same concept. In an other point of view, the system propose to create groups of conceptually similar images that may not be visually similar.

Consequently, it may be formalized as follows: if $G = (\mathcal{N}, \mathcal{E})$ is defined on the descriptions and images spaces $(\mathcal{D}, \mathcal{I})$, the description space is augmented to $\mathcal{D}_1$ by the user-defined descriptions and the dual functions r and r' are augmented to r_1 and r'_1 by the relations to the new descriptions. $\mathcal{I}$ remain unchanged.

It can also be seen as a set of new descriptions, each associated to a set of existing images.

4.2 Schemas Exploitation

In order to adapt the retrieval to the new space



$(\mathcal{D}_1, \mathcal{I})$, the system needs to adapt the graph structure presented to user. Moreover the changes must be applied immediately: user would not be satisfied by a system waiting for several semantic sets to be formed before adapting its structure. Consequently, we make use of an incremental algorithm. This algorithm aims at applying a subjective schema (i.e. a set of new descriptions and the associated images) to a Galois lattice.

We base our work on the algorithm already used for Galois' construction [14]; its complexity is experimentally $O(n)$ to add a element were n is the number of images in the base. However, this was measured in the case of adding one image and a set of description (actually, the complete set of descriptions associated to the image). In our case, since we are only adding a couple of description and one image we can expect better results.

From the user point of view, this system allows to define several subjective schemas that can be combined to form complex custom lattices. We can imagine users exchanging there own schemas on a given data set if they want to achieve similar goals.

5. Conclusion

In this paper, we proposed a new metadata classification separating objective ones from subjective ones, and combining external annotations and content-based metadata. The result is a navigation structure offering to the user the possibility to adapt itself according user's needs.

Compared to an approach proposing a before-hand annotation, our system doesn't suffer the gap between real user needs and what an expert think user will expect. Moreover, instead of requiring a costly annotation process, in our approach the human cost of the annotation is painlessly supported by user himself.

Our proposal also introduce a separation between the *objective* annotations that may not depend on the annotator and *subjective* annotations that will depend on a given user or a given application. Objective annotations may even be automatically retrieved if the acquisition source allows it.

This kind of retrieval system is of course irrelevant for an open world such as the world wide web, but is close to be a usable system in a private image collection, for example the catalog of an images provider. In this case a before-hand annotation is feasible and asking user an annotation on a base that will receive only few updates makes sense.

Galois lattices have an important drawback, we observe an explosion of the number of nodes while increasing the size of the base. Current proposal only aims at increasing relevance and do not answer to this problem.

In order to solves these issues related to Galois' lattices, it would be interesting to combine current metamodel with lattices masking, a technique to hide nodes and edges considered as irrelevant by user. However, both ask for user interaction and the way to merge user interfaces extensions introduced by these two techniques should be studied carefully; user should not be confused by providing similar feedback in two different systems.

文　　献

[1] V. Balasubramanian and M. Turoff. A systematic approach to user interface design for hypertext systems. In *Proceedings of the* 28^{th} *Hawaii International Conference on System Sciences (HICSS'95*, volume III, pages 241–250. IEEE Computer Science Press, 1995.

[2] F. Barbeau and J. Martinez. About tours in the OTHY hypermedia design. In Springer-Verlag, editor, *Proceedings of the* 5^{th} *International Computer Science Conference: Internet Applications (ICSC'99)*, number 1749 in Lecture Notes in Computer Science (LNCS), pages 146–155, Hong Kong, China, December 13-15 1999.

[3] S. Berchtold, D. A. Keim, and H.-P. Kriegel. The X-tree: An index structure for high-dimensional data. In *Proceedings of the* 22^{nd} *International Conference on Very Large Data Bases (VLDB'96)*, pages 28–39, 1996. Mumbai (Bombay), India.

[4] E. Bertino, B. C. Ooi, R. Sacks-Davis K.-L. Tan, J. Zobel, B. Shidlovsky, and B. Catania. *Indexing Techniques for Advanced Database Systems*. Kluwer Academic, Boston, Massachussets, 1997. 250 p.

[5] B. Bouchon-Meunier. *La logique floue et ses applications*. Addison-Wesley, October 1995. (in french), 257 p.

[6] C. Carson, S. Belongie, H. Greenspan, and J. Malik. Region-based image querying. In *Proceedings of the IEEE Worshop on Content-Based Access of Image and Video Libraries (CVPR'97)*, 1997.

[7] C. Carson and V. E. Ogle. Storage and retrieval of feature data for a very large online image collection. In *IEEE Computer Society Bulletin of the Technical Committee on Data Engineering*, volume 19, pages 19–27, December 1996.

[8] J. M. Corridoni, A. Del Bimbo, and E. Vicario. Image retrieval by color semantics with incomplete knowledge. *Journal of the American Society for Information Science*, 49(3):267–282, 1998.

[9] M. Corridoni, A. Del Bimbo, and P. Pala. Image retrieval by color semantics. *Multimedia Systems*, 7(3):175–183, 1999. Springer-Verlag.

[10] D. Dubois and H. Prade. Fuzzy sets in data summaries - outline of a new approach. In *Proc. of the 8th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'2000)*, volume 2, pages 1035–1040, Madrid, July 3-7 2000.

[11] M. Flickner, H. Sawhnery, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker. Query by image and video content: The QBIC system. *IEEE Computer*, 28(9):23–32, September 1995.

[12] T. Gevers. *Colour Image Invariant Segmentation and Retrieval*. PhD thesis, University of Amsterdam, The Netherlands, May 1996. 142 p.

[13] T. Gevers and A. Smeulders. A comparative study of several color models for color image invariant retreival, 1996.

[14] R. Godin, R. Missaoui, and H. Alaoui. Incremental concept formation algorithms based on Galois (concept) lattices. *Computational Intelligence*, 11(2):246–267, 1995.

[15] J. M. Kasson and W. Plouffe. An analysis of selected computer interchange color spaces. *ACM Transactions on Graphics*, 11(4):373–405, October 1992.

[16] Yves Lafon and Bert Bos. Describing and retrieving photos using rdf and http. http://www.w3.org/TR/photo-rdf/.

[17] Erwan Loisant, Hiroshi Ishikawa, Jose Martinez, Ohta Manabu, and Kaoru Katayama. User-adaptative navigation structures for images retrieval. In *Data Engineering Workshop (DEWS'2004)*, 2004.

[18] W. Y. Ma and B. S. Manjunath. NETRA: A toolbox for navigating large image databases. In *Proceedings of the IEEE International Conference on Image Processing (ICIP'97)*, 1997.

[19] J. Martinez and E. Loisant. Browsing image databases with Galois' lattices. In *Proceedings of the ACM International Symposium on Applied Computing (SAC'02)*, pages 971–975, Madrid, Spain, March 11-14 2002. ACM Computer Press.

[20] J. Martinez and S. Marchand. Towards intelligent retrieval in image databases. In B. Berra, editor, *Proceedings of the* 5^{th} *IEEE International Workshop on Multi-Media Data Base Management Systems (MMDBMS'98)*, pages 38–45, Dayton, Ohio, August 5-7 1998. IEEE Computer Press.

[21] C. Nastar, M. Mitschke, C. Meilhac, and N. Boujemaa. SurfImage: A flexible content-based image retrieval system. In *Proceedings of the* 6^{th} *ACM International Multimedia Conference (ACM-MM'98)*, Bristol, UK, September 14-16 1998. ACM, ACM Press.

[22] Yossi Rubner. Perceptual metrics for image database navigation. Technical Report CS-TR-99-1621, 1999.

[23] Simone Santini and Ramesh Jain. Integrated browsing and querying for image databases. *IEEE MultiMedia*, 7(3):26–39, – 2000.

[24] P. Valtchev, R. Missaoui, and P. Lebrun. A fast algorithm for building the hasse diagram of a Galois lattice. *Discrete Mathematics*, 2001.

[25] Y. J. Zhang. Evaluation and comparison of segmentation algorithms. *Pattern Recognition Letters*, 18:963–974, 1997.