Using ontologies to reduce the semantic gap between historians and image processing algorithms

Mickal Coustaty, Alain Bouju, Karell Bertet, Georges Louis
L3i labs
University of La Rochelle
La Rochelle, France
Email: {*mcoustat,abouju,kbertet,glouis*}*@univ-lr.fr*

Abstract—To reduce the gap between pixel data and the-saurus semantics, this paper presents a novel approach using mapping between two ontologies on images of drop-capitals (also named dropcaps or *lettrines*): In the first ontology, each dropcap image is endowed with semantic information describing its content. It is generated from a database of lettrines images - namely Ornamental Letter Images DataBase - manually populated by historians with dropcap images annotations. For the second ontology we have developed image processing algorithms to extract image regions on the basis of a number of features. These features, as well as spatial relations, among regions form the basis of the ontology. The ontologies are then enriched by inference rules to annotate some regions to automatically deduce their semantics. In this article, the method is presented together with preliminary experimental results and an illustrative example.

Keywords-Object recognition; Object segmentation; Knowledge based systems

I. INTRODUCTION AND CONTEXT

More and more historical documents are digitized to pre-serve them. This huge amount of data is still increasing and relevant tools are needed to retrieve them. In the particular case of graphical images, historians look for content-based image retrieval (CBIR) tools that will allow them to query large databases by the content of the images.

In this paper, we focus on CBIR techniques, applied to graphic images from documents of the XVth and XVIth, and more precisely on *lettrines* (see figure 1). A lettrine is a decorative letter at the beginning of a chapter or a paragraph in books of this period. These documents are provided by the CESR historical center[1]. Lettrines are binary stroke-based images, mainly containing a background (texture) and a letter. A CBIR tool, that allow to make query by example, would permit to search similar images, and to analyse theirs differences. Differences between similar images allow historians observing the degradations that appear as time passes.

Integration of semantic to the description of images would be relevant for this type of images from a high-knowledge domain. Usually, semantics are introduced as metadata de-fined by a thesaurus of the domain [IconClass, TEI, ...].

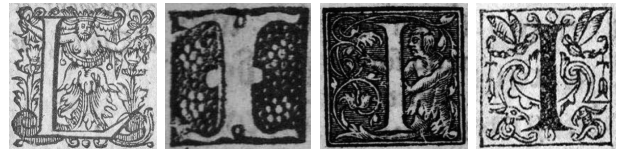[1]CESR historical center of the University of Tours

Figure 1. Samples of images of ancient ornamental letters - lettrines

Such thesaurus are used to manually annotate images, and are not integrated in classical CBIR tools which mainly focus on extracting information from images's content. The main challenge lies to an extraction of a semantic information automatically from images, in order to reduce the gap between pixels and semantics (*semantic gap*), similarly to [1], [2].

This paper introduces a novel ontology-based approach to automatically deduce semantic information from pixel data. Although the approach is applicable to graphical images in general, it is applied in this paper to lettrines.

Ontology technologies [3] have recently been introduced in the field of semantic web to organize data and knowledge of a domain in order to share, disseminate and update them. These technologies seem to provide a solid methodologi-cal framework, both relevant and innovative, to deal with document images. The ontology of lettrines proposed in this paper is obtained as a correspondence between two ontologies:

- In the first ontology (*ontology of historians' knowl-edge*), each lettrine image is described by metadata describing its content. It is generated from a database of lettrines images manually populated by historians using a specific thesaurus ;
- In the second ontology (*ontology of image process-ing*), each lettrine image is described by a set of regions. These regions correspond to the segmentation of principal shapes in the image, using automatic image processing treatment that we have developed [4], [5]. These regions are described by statistical informations and by spatial relations between them.

The ontology we have developed combines the knowledge of historians as well as information about the regions of the

image, and can be queried on both type of properties.

Description logic inference rules are used for deducting knowledge that is implicit in the knowledge-base. Indeed, some low-level features of a region (area, eccentricity, color, ...), and its spatial position (center of the image, near the sides, ...) can help to characterize its content. The expression of this characterization allows to envisage a mechanism of automatic annotation of semantics of a region using low-level features. We conducted some experiments in this direction.

Section 1 defines the ontology of historian knowledge ; Section 2 describes the ontology of image processing and spatial relations ; Sections 3 presents the final ontology, where inference rules are used to infer new knowledge.

II. Ontology of Historian knowledge

In historical area research, lettrines are classically split into three layers [6] (see Figure 2). Each layer capturing one kind of knowledge about the lettrine:

- Letter: at the center of the lettrine, the letter can be black or white, and with different fonts ;
- Pattern: it is composed of all ornamental form either decorative (meaningless drawings) or figurative (people, nature, building ...) ;
- Background: it can be either uniform (black or white), dashed or streaked ;
- Frame: it defines the boundaries of the typographical block, composed of zero, one or two strokes.

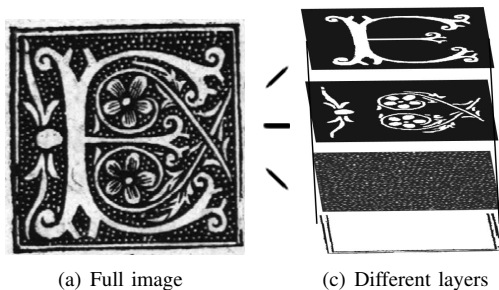(a) Full image   (c) Different layers

Figure 2.   Structural approach of historians for a Lettrine

Historians of the CESR historical center of University of Tours have defined and manually populated a thesaurus for 4288 images of lettrines. Their thesaurus respects the Text Encoding Initiative [3]and the Iconclass classification [7].

From the properties of Images as identified by historians, it is easy to construct the T-Box (terminology box) of an ontology. Concepts (vocabulary of description logic) or classes (vocabulary of OWL) were defined for every property. Here is the resulting classes hierarchy:

- Lettrine (Image of)
- Background
  - BlackBackground

- WhiteBackGround
  - DashedBackground
  - StreakedBackground
  - UndefinedBackground (unable to classify)
- Pattern
  - FigurativePattern
  - DecorativePattern
- Alphabet
  - GreekAlphabet
  - HebrewAlphabet
  - LatinAlphabet
- TypeFont (only applies to latin letters)
  - RomanFont
  - GothicFont
  - UndefinedFont (All other cases)
- Letter (A, B, C, ..., Z, Γ, Δ, ..., Ω, $\aleph$, ...)
- Frame
  - NoFrame
  - FrameMat (a wooden or metalic frame where the lead typographic character is inserted)
  - Framing
  - Strokes
    * OneStroke
    * TwoStrokes
  - OtherFrame

Object properties are then define among those classes :

- Lettrine *hasBackground* Background
- Lettrine *hasPattern* Pattern
- Lettrine *hasAlphabet* Alphabet
- Lettrine *hasTypeFont* TypeFont
- Lettrine *hasLetter* Letter
- Lettrine *hasFrame* Frame

The information about the 4288 individual lettrines indexed by historians was used to populate the corresponding A-Box (assertion box). The ultimate goal is to perform automatic classification of lettrines into that A-Box. But the ontology also serves as a guide when we need to retrieve lettrines for image processing experiments.

III. Ontology of image processing and spatial relations

In this section, and for the remaining of the paper, we will use the terms shapes or region to qualify shapes automatically extracted in our images.

A. Region extraction

Images of lettrines, being composed of varied superposed layers, represent a rich but complex source of information. Thus, extracting that information from these images is particularly difficult. The letter and the shapes suffer from many degradations due to the printing substrate and the period of creation of these documents.

These images are mainly made up of lines, which are unsuitable for the usual texture methods. We have developed an approach in [4] to simplify and to extract shapes from images with lines. The process of shapes extraction thus relies on a three step process:

1) Firstly, the lettrine's images are simplified using a decomposition which relies on a series of projection (see [8], [9] for complete explanations). An example of result can be observed in figure 3). The first layer obtained by the decomposition enables to highlight geometry, which corresponds to shapes in the image.

2) Once this first step has been performed, all the connected component of this layer are segmented using a Zipf law [10]. This statistical law present the advantage to be robust toward the variation of greylevel and is independent from color of shapes ;

3) Finally, all the connected components are extracted using the algorithm defined in [11]. Among all these connected components, we only retain shapes whose area is superior to 1% of the image (shapes that are composed with less than 1% of pixels are not relevant for image description).

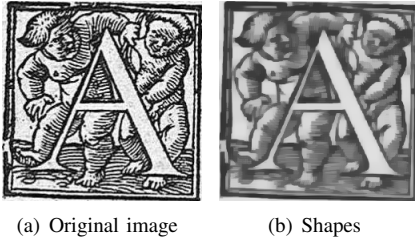   (a) Original image   (b) Shapes

Figure 3.   An example of drop cap decomposition using decomposition presented in [4]

Each extracted shape is described by a vector of classical features:

$$(G_C;\ Area;\ E_n;\ P_{feat}) \qquad (1)$$

where :

G_C : indicates the coordinates of the center of gravity, allowing to locate shape in the image ;

$Area$: gives information about the size of shapes (to sort them for example) ;

E_n : corresponds to the Euler Number [12]. In our case, since there is a single region, the Euler Number is reduced to:

$$E_n = 1 - H \qquad (2)$$

where H represents the number of holes.

Figure 4.   Partition of the images used to locate the regions in the center of close to the side.

P_{feat} : corresponds to a list of nine boolean values that indicates if the shape intersects with subparts of the image. These subparts result from a division in nine fixed parts of the image (see figure 4 - subparts are listed from the number 1 to the number 9). Part number 5 corresponds to the center of the image, while the others correspond to the side parts.

The partition features are used to locate a shape in an image, and more precisely to detect if a shape is located in the center of the image, or if it is connected with the side of the image. For example, a shape corresponding to a letter has to be located in the center of the image without connection with any side.

B. Ontology

An ontology of image processing is used to produce a knowledge-base of extracted regions. For better interoperability, extracted shapes are represented by concepts using the standard Simple Feature Interface Standard (SFS) developed by the Open GIS Consortium [13]. Here is the resulting class hierarchy:

- **Image (of lettrine):** image itself;
- **ImageMeasurementUnit:** to precise the mesure unit for image size;
- **ImageReferenceSystem:** the reference system of the image uses to indicate coordinate of pixels;
- **SurFace:** the main abstract class of SFS, derived in;
  - **ImagePartition:** correspond to the subpart of the grid presented in figure 4;
  - **Region:** each shape extracted using our algorithm;
- **Centroid:** indicated the coordinates of the center of gravity of each shape
- **RegionSet:** a set of shapes extracted from an image;

The standard SFS used in this model allows computing spatial relation between objects using the RCC-8 algebra. In order to be able to compute spatial relation between the shapes and the grid defined in figure 4, the concepts *Region* and *ImagePartition* inherit from the concept SurFace defined in the standard SFS.

Ontology is enriched with properties of images, properties of regions, and spatial properties between regions and image

part. The first kind of property allow extracting information about the image itself:

- *Image **hasImageMeasurementUnit** MeasurementUnit,* unit used in the different measures obtained from image processing;
- *Image **hasImageReferenceSystem** ImageReferenceSystem,* the reference system used;
- *Image **hasRegionSet** RegionSet,* information about the set of region of an image;
- *Image **hasPartition** Partition,* information about the subpart of the grid defined above;
- *Image **hasHeight**,* and *Image **hasWidth**,* dimensions size of the image;
- *RegionSet **hasRegion** Region,* information about the region of an image.

In a second time, a set of properties to manage knowledge about information extracted from each region:

- *Region **hasCentroid** Centroid,* the coordinates of the centroid of each shape;
- *Region **hasX*** and *Region **hasY**,* to retrieve the X,Y coordinates of the center of gravity;
- *Region **hasAire**,* to retrieve the area of the shape in pixel;
- *Region **hasEuler**,* to retrieve the number of holes in the shape.

Spatial relations are used to extract information about the position of a shape in the grid defined above. These properties are inherited from the RCC-8 algebra (meaning of the relations can be found in [14]).

- *Region **disconnectedFrom (DC)** Part*
- *Region **externallyConnectedTo (EC)** Part*
- *Region **partiallyOverlaps (PO)** Part*
- *Region **equalsTo (EQ)** Part*
- *Region **tangentialProperPartOf (TPP)** Part*
- *Region **nonTangentialProperPartOf (NTPP)** Part*
- *Region **hasTangentialProperPart (TPPi)** Part*
- *Region **hasNonTangentialProperPart (NTPPi)** Part*

Image processing were used on the database of lettrines. We selected a subset of 100 images to populate our knowledge-base. From these images, we obtained a set of 584 regions.

IV. Final ontology and inference rules

We construct a correspondence between the ontology of historians' knowledge, and the ontology of image processing and spatial relations. Semantics informations are deduced from low-level features of the image using inference rules. Inference rules are applied on the correspondence between the ontology of historians' knowledge and the ontology of image processing and spatial relations.

A. Final ontology

Since the LETTRINE/IMAGE concept appears in each ontology, it is possible to link these concepts in order to
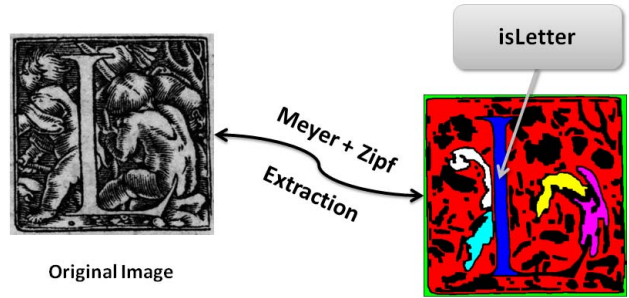

Figure 5. Shapes extraction in lettrines inspired from [4], and their label attributed using the knowledge-base of the proposed approach

combine the dropcap knowledge from the two ontologies. We use the OWL constructor OWL:EQUIVALENTCLASS (OWL first version[2]) to link these two concepts (T-Box) and the constructor OWL:SAMEAS for the A-Box link. The knowledge database is populated with all data.

The first ontology is populated with 4288 individual lettrines, and the second one with 589 regions extracted for 100 images. The mapping of the two ontologies gives raise to a knowledge base. In order to extract information from these database, we defined rules to populate new properties of regions. These properties correspond to an automatic semantic annotation of shapes in images.

B. Rules and Experiments

The two ontologies - ontology of historian knowledge ; ontology of image processing and spatial relations - are created (T-Box) using Protégé 3.4.4 and populated (A-Box) using Jena[3].

Using Jess rule engine, rules inference is applied on extracted regions to automatically annotate the semantics of some of them. Rules we defined are used to populate new properties of regions. Here, we present an example of property and rule that detect the letter:

- *Region **isLetter***: indicates that the region is identified as the letter of the lettrine;

The rule for the *isLetter* property is dedicated to annotate regions corresponding to the letter of lettrine. We establish that a region corresponds to the letter when it is:

(isLetter) *the maximal area region located in the center of the image with few holes.*

More precisely, the *isLetter* property is expressed in SWRL as a combination of the following properties:

- **Located in the center area:** This property is defined by testing if the region is contained in the center partition

(part number 5) and not overlapped by sides (other parts);

- **Region with few holes:** The Euler number is used to test this property. By Some experiments, we position the Euler number between -2 and 2 to characterize a region without holes.
- **Maximal area region:** This property is the last one to be tested. Among regions verifying the first two properties, the one having the maximal area is finaly annotated with the *isLetter* property.

Applied on a subset of 100 images of lettrines, we obtained a set of 584 regions. Among these regions, 112 are located in the center, and 100 are labelled with the *isLetter* property (only one shape by image can be labelled with this property). A manual verification allows to state that 91 regions are well-annotated, and we thus obtained a recognition rate of 91%.

V. CONCLUSION AND PERSPECTIVE

This work is a contribution to the semantic gap problem between pixels and semantics. It focuses on the automatic extraction of semantic informations from images. The proposed approach is a modular combination of informations issued from automatically extracted regions of historical images, and from thesaurus of the historians' domain. This combination is used to automatically deduce the semantic of some extracted regions.

In this paper, we describe an ontology for images of lettrines resulting from a mapping between an ontology of historian knowledge and an ontology of image processing and spatial relations. We extend these ontologies using inference rules to automatically annotate some regions. The presented rule allow to retrieve the shape associated to the letter. This rule is only defined from low-level features of regions. We present an experiment to illustrate the efficiency of this rule.

We are presently working on labelling shapes that correspond to body part of the character in the background of lettrines.

Another extension would consist in comparing the shape of such annotated regions with the expected shape of the letter. Indeed, the identification of the letter of the lettrine, information issued from the ontology historian knowledge, could enable to precise the expected shape. Such an extension could allow to detect the special case when a letter is decomposed in several regions, and thus to reiterate the *isLetter* property until all regions of the letter are detected.

An improvement would be to define new regions rules for other semantic annotation. Those rules could use a more complex image partition, other image processing features, and information issued from the thesaurus. Therefore, even if the application is dedicated to lettrines, our goal is to use the "image processing and spatial relations" ontology with other kind of graphical image.

REFERENCES

[1] V. Mezaris, I. Kompatsiaris, and M. G. Strintzis, "Region-based image retrieval using an object ontology and relevance feedback," *EURASIP J. Appl. Signal Process.*, vol. 2004, pp. 886–901, January 2004.

[2] K.-W. Park, J.-W. Jeong, and D.-H. Lee.

[3] M. Uschold and M. King, "Ontologies: principles, methods and applications," *Knowledge Engineering Review*, vol. 11, no. 2, pp. 93–155, 1996.

[4] M. Coustaty, J.-M. Ogier, R. Pareti, and N. Vincent, "Drop Caps Decomposition For Indexing - A New Letter Extraction Method," in *10th International Conference on Document Analysis and Recognition*, Barcelona Espagne, 2009, pp. 476–480, IAPR.

[5] S. Jouili, M. Coustaty, S. Tabbone, and J.-M. Ogier, "Navidomass: Structural-based approaches towards handling historical documents," in *ICPR*. IEEE, 2010, pp. 946–949.

[6] R. Jimenes, "Les bibliothèques virtuelles humanistes et l'étude du matériel typographique," Centre d'Etude Superieur de la Renaissance, Tech. Rep., 2008. [Online]. Available: http://www.bvh.univ-tours.fr/Note_typo.pdf

[7] H. v. d. Waal, *ICONCLASS, an Iconographic Classification System*. Completed and edited by L. D. Couprie et al, 1985. [Online]. Available: http://www.iconclass.nl/

[8] S. Dubois, M. Lugiez, R. Péteri, and M. Ménard, "Adding a noise component to a color decomposition model for improving color texture extraction," in *4th ECCGIV*, 2008, pp. 394–398.

[9] A. E. Hamidi, M. Menard, M. Lugiez, and C. Ghannam, "Weighted and extended total variation for image restoration and decomposition," *Pattern Recognition*, vol. 43, no. 4, pp. 1564 – 1576, 2010.

[10] R. Pareti and N. Vincent, "Ancient initial letters indexing," in *18th International Conference on Pattern Recognition*. Hong Kong, China: IEEE Computer Society, 2006, pp. 756–759.

[11] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*, 1st ed. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1992.

[12] W. K. Pratt, *Digital Image Processing: PIKS Scientific Inside*, 4th ed. Wiley-Interscience, Feb. 2007.

[13] M. F. Goodchild, *Interoperating geographic information systems*. Springer, Feb. 1999.

[14] D. A. Randell, Z. Cui, and A. G. Cohn, "A spatial logic based on regions and connection." in *KR'92*, 1992, pp. 165–176.