Automatic indexing of comic page images for query by example based focused content retrieval

Muhammad Muzzamil Luqman, Hoang Nam Ho, Jean-Christophe Burie, Jean-Marc Ogier

Automatic indexing of comic page images for query by example based focused content retrieval

Muhammad Muzzamil Luqman, Hoang Nam Ho, Jean-Christophe Burie and Jean-Marc Ogier
L3i Laboratory, University of La Rochelle. 17042 La Rochelle FRANCE
{muhammad_muzzamil.luqman,hoang_nam.ho,jean-christophe.burie,
jean-marc.ogier}@univ-lr.fr

***Abstract*—Automatic indexing of comic-page-image repositories has evolved as an interesting research problem for graphics recognition research community. In this paper we present a system for automatically indexing the comic-page-images in order to achieve query by example (QBE) based focused content retrieval. In our system we represent the comic-page-images by attributed graphs and translate the problem of *automatic indexing / QBE based focused content retrieval* as a subgraph spotting problem. Our system uses an explicit graph embedding technique to embed the comic-page-image graphs into numeric feature vectors and then employs state-of-the-art machine learning tools for *automatic indexing / QBE based focused content retrieval*. Experimental results are presented for automatic indexing and QBE based focused content retrieval in a comic-page-image repository.**

I. Introduction

In recent years the automatic indexing and content based focused retrieval in comic-page-image repositories, has emerged as an interesting research problem for graphics recognition research community. Comics represent an important heritage in many countries. The digitization of paper based comic archives give rise to interesting research challenges. Considering the increasing size of comic-page-image repositories, unsupervised algorithms are desired for automatic indexing, so that the liberty of natural ways of querying the graphics content in these repositories (i.e. query by example (QBE) based focused retrieval, instead of traditional text based information retrieval) could be provided to the users.

The comic pages are rich in structured graphics content and in our point of view the research problem of content retrieval in comic pages provides an opportunity to apply existing methodologies of graphics recognition. Graph based representations are the natural, most powerful and widely used data structures for structured data [1]. Graphs provide unmatched advantages for representing the topological and geometric details of underlying data but graphs have limited computational efficient tools to process them. Modern approaches of embedding graphs in vector spaces allow to overcome the computational expensiveness of graphs [2][3] and enable graph based indexing and retrieval of image repositories.

An interesting recent work on (Japanese comic) manga retrieval is presented in [4]. The authors have used bag-of-features method using visual words based on face ROIs and clusters of generic ROIs, for similar manga retrieval. We believe that graph based representation of comic pages provides a natural way to perform many high level semantic operations in comic-page-image repositories. Our current work is inspired by the work in [4] and we have investigated a graph based methodology for comic retrieval.

In this paper we present a system for automatic indexing and QBE based focused content retrieval in comic-page-image repositories. The block diagram of our system is presented in Figure 1. In our system the comic pages are represented by region adjacency graphs (RAGs) - automatic and unsupervised indexing is achieved by employing graph embedding and machine learning tools - and QBE based focused content retrieval is achieved through subgraph spotting. The work presented in this paper is an advancement of our research on subgraph spotting [5]. The novel contributions of this paper are the RAG based graph representation for comic pages, adaptation of our subgraph spotting system to comic page graphs and experimental evaluation of the new system for comic-page-image repository.

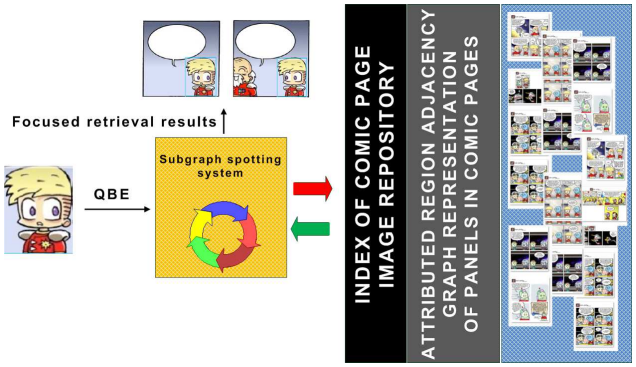


Fig. 1. Block diagram of proposed system for comic retrieval.

In Section II we introduce the notations and definitions of the concepts used in this paper. Section III details the graph representation phase of our system, the automatic indexing phase and QBE based focused content retrieval in comic-page-image repositories. Experimental results and discussion are presented in Section IV. The paper concludes in Section V with futures lines of work.

II. Definitions and notations

Attributed Graph (AG): Let A_V and A_E denote the domains of possible values for attributed vertices and edges. These domains are assumed to include a special value that represents a null value of a vertex or an edge. An attributed graph AG over (A_V, A_E) is defined to be a four-tuple:

$$AG = (V, E, \mu^V, \mu^E)$$

where,
V is a set of vertices,
$E \subseteq V \times V$ is a set of edges,
$\mu^V : V \longrightarrow A_V$ is function assigning attributes to vertices and
$\mu^E : E \longrightarrow A_E$ is a function assigning attributes to edges.

In an attributed graph AG: $|V|$ is graph order, $|E|$ is graph size and degree of a node is number of edges connected to it.

Resemblance attributes: The resemblance attributes are new attributes that we compute from the attributes of nodes and edges in a graph. These attributes provide a measure of homogeneity among the surrounding nodes of an edge and vice versa.

An edge connects exactly two nodes in a graph. For an edge, the resemblance attributes are computed from respective attributes of its nodes. For example, for an edge between nodes, "1" and "2", resemblance between a numeric attribute "a" is computed by Equation 1 and resemblance between a symbolic attribute "b" is computed by Equation 2.

$$resemblance = min(|a_1|, |a_2|)/max(|a_1|, |a_2|) \quad (1)$$

$$resemblance = \begin{vmatrix} 1 & b_1 = b_2 \\ 0 & otherwise \end{vmatrix} \quad (2)$$

A node can have many edges connecting it to other nodes in the graph. For a node, the resemblance attributes are computed as the mean of the resemblance between all pairs of its edges. The latter is computed by Equation 1 for a numeric attribute "a" and by Equation 2 for a symbolic attribute "b".

Graph Embedding: Graph Embedding is a methodology aimed at representing a whole graph (with attributes attached to its nodes and edges) as a point in a suitable vector space; preserving the similarity of the graphs i.e. the more two graphs are considered similar, the closer should be the corresponding points in the vector space. The graph embedding methods are formally categorized as implicit and explicit graph embedding. The implicit graph embedding methods are based on graph kernels, and actually are functions that perform scalar product between the vectors associated to two graphs. Such an embedding satisfies the main mathematical properties of scalar product but does not permit all the operations that could be defined on vector spaces. On the other hand, the more useful, explicit graph embedding methods map a graph to a point in suitable vector space. It encodes the graphs by equal size vectors and produces one vector per graph; thus enabling the use of all the methodologies, techniques and operations devised for vector spaces. The vectors for two graphs which are obtained by an explicit graph embedding method can also be employed in a standard dot product for defining an implicit graph embedding function between them.

Fuzzy Multilevel Graph Embedding (FMGE): Fuzzy Multilevel Graph Embedding (FMGE) is a method of explicit graph embedding, which has been proposed in [6]. FMGE exploits structural and statistical significant details of an attributed graph for embedding it into a feature vector. Its resulting feature vector is termed as Fuzzy Structural Multilevel Feature Vector (FSMFV). As shown in Figure 2 FSMFV contains graph level features, node level features and edge level features. The graph level features are graph order and graph size. Node level features are fuzzy histogram of node degree and fuzzy histogram of each distribution of the values taken by node attributes. And the edge level features are fuzzy histogram of each distribution of the values taken by edge attributes. FMGE employs fuzzy overlapping trapezoidal intervals for minimizing the information loss while mapping from continuous graph space to discrete vector space. FMGE permits graphs to benefit from state of the art computational efficient, clustering and classification tools, which originally is not possible in graph space.

Graph retrieval and subgraph spotting: Graph retrieval deals with retrieving a graph G from a graph repository, based on the similarity of graph G with an example (or query) graph.

Subgraph spotting takes the definition of graph retrieval to further granularity. Subgraph spotting refers to retrieving graph G from graph repository based on the similarity of a subgraph in G with the example (or query) graph.

III. Automatic indexing and QBE based focused content retrieval system

This section details our proposed system for automatic indexing and QBE based focused content retrieval in comic-page-image repositories. In Section III-A we discuss the graph representation phase of comic pages and in Section III-B we outline the indexing/retrieval phase of the system.

A. Graph representation phase

This section describes our approach for representing comic-page-images by graphs. We first extract the regions of interest in the comic-page-image and then construct an attributed region adjacency graph (RAG) for each panel in the comic-page-image.

Preprocessing comics page images: A comic-page-image contains several panels. In order to represent each panel by a RAG, we first extract the panels in the comic-page-image. For this step we employ the method proposed by Rigaud et al. in [7]. The method of Rigaud et al. exploits surrounding black lines of panels and text balloons in comic-page-images and uses a connected-component labeling analysis for automatically extracting the panels and text.

After splitting a page into panels, in the next step we segment each panel into regions according to selected criteria as detailed in Figure 3. The segmentation of panels into regions is a very important step and it has a great influence on construction of graphs.

In Figure 3 filtering regions refers to eliminating areas that are not important. The segmentation step creates lots of regions in panels. The black lines delimiting the region's color are not essential in our approach. A lot of small regions are also created by the segmentation step. These small regions are not useful and must be ignored for limiting the size of the

Graph order	Graph size	Embedding of node degree	Embedding(s) of node attribute(s)	Embedding(s) of edge attribute(s)

Fig. 2. Feature vector of FMGE. The feature vector is comprised of *exactly one* feature for each of graph order and graph size, features for node degree, features for node attributes and features for edge attributes.

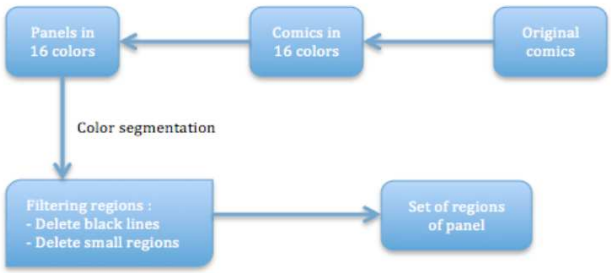


Fig. 3. Pre-processing comics.

resulting graph. We have defined the criteria of compactness for calculating the proportion between the surface of the region and the bounding box that contains it. If the proportion is less than a certain threshold we ignore the region.

The preprocessing step gives us a set of regions in a panel of the comic-page-image. Each region is characterized by a list of attributes including the color, the surface and the compactness of the region. A list detailing the adjacent regions is also maintained by each region. Figure 4 provides an example for the preprocessing step of comic-page-images.

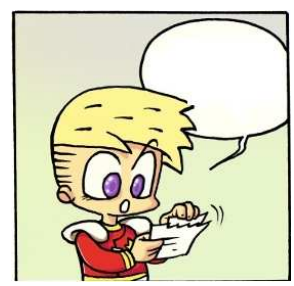
Graph construction: The extracted regions of a panel form the list of nodes of the graph i.e. each region is represented as a node in the graph. The nodes are labeled by color, compactness and shape attributes. The color is defined in CIE-Lab color space. This color space has been designed to approximate perceptually uniform color spaces. Each region is described by a vector of three dimensions where each element corresponds respectively to the values of L (Luminance), a and b (two colors channels). A second attribute associated to the nodes is the compactness of the region. This value is calculated during the preprocessing step. The shape of the region represented by a node is a very important discriminatory characteristic. We use Hu moments because of their ease of computation and robustness to rotation, translation and scaling. These properties are necessary in our case because the comics contains many complex objects (such as the comic characters) which can be of different size (close or distant view) and can be drawn in different orientation. We extract seven values of Hu moment for each region.

Thus a node in the graph (which is representing a region in the panel) is attributed by a vector of 11 attributes. These attributes include three attributes for color, one for compactness and seven Hu moments for shape of the underlying region. Each of the attributes are normalized separately. Suppose that x_i is an observed value of an attribute, $\bar{x}$ the average value and Γ the standard deviation of the attribute. The normalization is performed as follows: $(x_i - \bar{x})$ / Γ.

The edges of the graph are obtained by the spatial relationships between the regions represented by nodes of the graph. This allows us to construct a region adjacency graph

(a) Comic-page-image

(b) Panel extraction



(c) Preprocessing

Fig. 4. (a)A comic page. (b) An extracted panel. (c) Panel after preprocessing.

(RAG) which incorporates the topological relations between the regions of the comic-page-image panel.

Among the spatial relations given in RCC-8 [8], the only ones that are relevant in the case of comics are the adjacency and inclusion properties. If we consider the drawings as a set of regions, a region can only be adjacent or included in another. In our case, we use only the adjacency relation between regions to define the edges of the graph.

The label of the edges is defined from the surface area ratio of adjacent regions. We assume that the proportions between regions are preserved whether they are viewed from near or far. The ratio between the surfaces of two adjacent regions is invariant to scale.

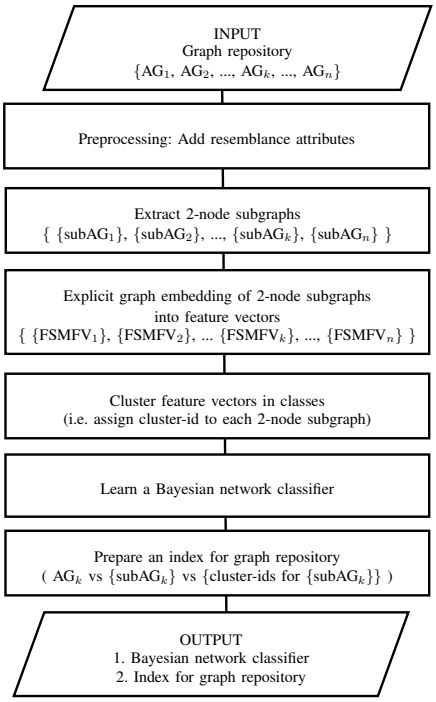


Fig. 5. Automatic indexation of a graph repository.

B. Indexing and subgraph spotting phase

The system outlined in this section is reproduced from our recent work in [5][6]. We achieve graph retrieval and subgraph spotting through explicit graph embedding.

As outlined in Figure 5, in our system a graph repository is automatically indexed during an off-line learning phase; where we (i) compute resemblance attributes for each node and edge in graph, (ii) break the graphs into 2-node subgraphs (a.k.a. cliques of order 2), which are primitive building-blocks of a graph, (iii) embed the 2-node subgraphs into feature vectors by employing our recently proposed explicit graph embedding technique, (iv) cluster the feature vectors in classes by employing a classic agglomerative clustering technique, (v) build an index for the graph repository and (vi) learn a Bayesian network classifier.

The subgraph spotting is achieved during the on-line querying phase; where we (i) compute resemblance attributes for each node and edge in graph, (ii) break the query graph into 2-node subgraphs, (iii) embed them into feature vectors, (iv) employ the Bayesian network classifier for classifying the query 2-node subgraphs and (v) retrieve the respective graphs by looking-up in the index of the graph repository. The graphs containing all query 2-node subgraphs are retrieved to form the set of result graphs (AG_{result}) for the query (AG_{query}).

For spotting the query graph AG_{query} in a result graph AG_{result}, we employ the adjacency matrix of graph AG_{result} along-with a score function. The adjacency matrix of graph AG_{result} has a value of "0" if there is no edge between "$node_1$" and "$node_2$" in the original graph AG_{result}, a value of "1" if there is an edge between "$node_1$" and "$node_2$" in the original graph AG_{result} and a value of "2" if one of the query 2-node subgraphs is classified (by Bayesian network) as belonging to the cluster of this 2-node subgraph (which is comprising of edge between "$node_1$" and "$node_2$"). The query graph AG_{query} is finally spotted in the result graph AG_{result}, by looking up in the neighborhood of each 2-node subgraph of AG_{result} which is in the result i.e. each "$AG_{result}(i,j) = 2$" in the adjacency matrix of result graph AG_{result}. We explore "o" connected neighbors of each "$AG_{result}(i,j) = 2$". The parameter "o" is proportional to the graph order of query graph AG_{query} ($|V_{query}|$). We compute a score for each "$AG_{result}(i,j) = 2$" using Equation 3. The computed score of "$AG_{result}(i,j) = 2$" also gives a confidence value for subgraph spotting of query graph AG_{query} in result graph AG_{result}.

$$score = \sum_{z=0}^{2} (z \times \frac{|z|}{o}) \tag{3}$$

where ,
- z is a value in the adjacency matrix (either 0,1 or 2),
- $|z|$ is number of times value z occurs in neighborhood
- o is number of connected neighbors that are looked-up.

IV. Experimentation

We have evaluated the empirical performance of our proposed system on a comic-page-image. The repository was comprised of a total of 216 panels extracted from 94 comic-page-images.

As outlined in Section III-A for each of the panel image, preprocessing step was performed to extract the distinct regions in the image. These regions were employed to construct an attributed region adjacency graph (RAG) representation for the panel from comic-page-image.

After representing the panels from comic-page-image by attributed region adjacency graphs (RAGs), we employ our system outlined in Section III-B for automatically indexing these graphs. During the off-line indexing phase, first new resemblance attributes are computed from the corresponding attributes of the nodes and edges in the graph for encoding the homogeneity of the neighborhood of the nodes and edges respectively. The cliques of order-2 are extracted from the graphs and are embedded into feature vectors by employing FMGE. The feature vectors are embedded into classes by hierarchical clustering. The clusters are used as classes for learning a Bayesian network classifier and an index is built for the RAG representation of the comic-page-image repository. This index maps the comic-page-image panels $\longrightarrow$ their graphs representations $\longrightarrow$ cliques of order-2 $\longrightarrow$ feature vector embedding of cliques.

The QBE based focused content retrieval is achieved by querying the system with a selected part of a comic-page-image panel. The system represents the query image by a region adjacency graph, computes the resemblance attributes for nodes and edges, extracts order-2 cliques from the graph, embeds the cliques by FMGE feature vector and employs the Bayesian network classifier for classifying each of the cliques in the query graph. The panels containing all of the cliques of the query graph are retrieved as results of the query. Finally we

employ the score function presented in Equation 3 for focusing on the interesting part of the retrieved panel.

We have tested our system for 25 query images. Figure 6 presents the precision and recall plot for the retrieval results of our system. For a 100% recall the system achieves a precision of approximately 50%. The preliminary experimentation has been performed on a limited set of comic-page-images but results are encouraging and illustrates the applicability of the proposed approach for comic-page-image indexing and QBE based focused retrieval. Given the fact that due to propriety issues there is a lack of public comic-page-image repositories, we have experimented with a home-available moderate size comic-page-image repository.

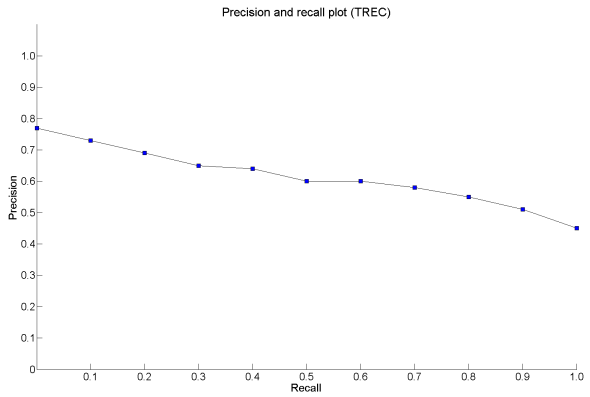


Fig. 6. Precision and recall plot for comic-page-image retrieval.

V. Conclusion

We have presented a system for automatic indexing of comic-page-image repository for query by example based focused content retrieval. We have modeled the problem of comic-page-image retrieval as a subgraph spotting problem. The system provides unsupervised indexing of comic-page-image repositories and offers the ease of query by example along-with the granularity of focused retrieval. Our preliminary experimentation results are very encouraging and in future we plan to take this work forward by investigating further into the application of existing graphics recognition techniques for comic-page-image indexing/retrieval. Our future direction of research is to take forward our work by exploring cliques of higher order ($\geq$3) for building a multi-resolution index of a graph repository. We also plan to evaluate the performance of our system on more comic-page-image repositories.

References

[1] D. Conte, P. Foggia, C. Sansone, and M. Vento, "Thirty years of graph matching in pattern recognition," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 18, no. 3, pp. 265–298, 2004.

[2] P. Foggia and M. Vento, "Graph Embedding for Pattern Recognition," in *Recognizing Patterns in Signals, Speech, Images and Videos*, ser. Lecture Notes in Computer Science, D. Ünay, Z. Çataltepe, and S. Aksoy, Eds. Springer Berlin Heidelberg, 2010, vol. 6388, pp. 75–82.

[3] H. Bunke and K. Riesen, "Recent advances in graph-based pattern recognition with applications in document analysis," *Pattern Recognition*, vol. 44, no. 5, pp. 1057–1067, May 2011.

[4] W. Sun and K. Kise, "Similar manga retrieval using visual vocabulary based on regions of interest," *International Conference on Document Analysis and Recognition*, pp. 1075–1079, 2011.

[5] M. M. Luqman, J.-Y. Ramel, J. Lladós, and T. Brouard, "Subgraph spotting through explicit graph embedding: An application to content spotting in graphic document images," in *ICDAR*, 2011, pp. 870–874.

[6] M. M. Luqman, J.-Y. Ramel, J. Lladós, and T. Brouard, "Fuzzy multilevel graph embedding," *Pattern Recognition*, vol. 46, no. 2, pp. 551 – 565, 2013.

[7] C. Rigaud, N. Tsopze, J.-C. Burie, and J.-M. Ogier, "Robust Frame and Text Extraction from Comic Books," in *Graphics Recognition. New Trends and Challenges*, ser. Lecture Notes in Computer Science, Y.-B. Kwon and J.-M. Ogier, Eds. Springer Berlin Heidelberg, 2013, vol. 7423, pp. 129–138.

[8] A. Cohn, B. Bennett, J. Gooday, and N. Gotts, "Representing and reasoning with qualitative spatial relations about regions," *Spatial and temporal reasoning*, pp. 97–134, 1997.