Spatial Relationships in Visual Graph Modeling for Image Categorization

Trong-Ton Pham
Grenoble INP-LIG
385 Av. de la Bibliothèque
Grenoble, France
ttpham@imag.fr

Philippe Mulhem
CNRS-LIG
385 Av. de la Bibliothèque
Grenoble, France
mulhem@imag.fr

Loïc Maisonnasse
TecKnowMetrix
4 rue Léon Béridot
Voiron, France
lm@tkm.fr

ABSTRACT

In this paper, a language model adapted to graph-based representation of image content is proposed and assessed. The full indexing and retrieval processes are evaluated on two different image corpora. We show that using the spatial relationships with graph model has a positive impact on the results of standard Language Model (LM) and outperforms the baseline built upon the current state-of-the-art Support Vector Machine (SVM) classification method.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Algorithms, Experimentation

Keywords

Graph theory, language model, image categorization

1. VISUAL GRAPH MODELING

Our goal here is to automatically induce, from a given image, a graph that represents the image content. This graph will contain concepts directly associated with the visual elements in the image, as well as relations which express how concepts are related spatially in the image. To do so, our procedure is based on four main steps: (1) Identify regions within the image that will form the basic blocks for concept identification; (2) Index each region with a predefined set of features; (3) Cluster all the regions found in the collection in k classes, where each class represents one concept. Each region in the image is then associated to one concept. We obtain for each type of region a set of concepts $\mathcal{C}$; (4) Finally, extract spatial relations between visual concepts.

Based on this representation, our contributions are twofold. First, a directed graph model is constructed to represent the image content based on concepts. Second, we apply a simple and effective method for the graph matching based on the language model [2]. Unlike the previous approach [3] based on a sequence of n-gram concepts, our framework embeds smoothly different types of spatial relations. The experiments carried out on two image collections confirm the significative impact of our method.

SIGIR'10, July 19–23, 2010, Geneva, Switzerland.
ACM 978-1-60558-896-4/10/07.

1.1 Graph definition

We assume that each image i is represented by a set of weighted concept sets $S^i_{WC} = \{W^i_C\}$ and a set of weighted relation sets $S^i_{WE} = \{W^i_E\}$, forming a graph:

$$G^i =< S^i_{WC}, S^i_{WE} >$$

Each concept of one set W^i_C corresponds to a visual concept used to represent the image according to the feature associated with. Denoting $\mathcal{C}$ a set of concepts for one feature over the whole collection, W^i_C is a set of pairs $(c, \#(c,i))$, where c is an element of $\mathcal{C}$ and $\#(c,i)$ is the number of times c occurs in the document image i:

$$W^i_C = \{(c, \#(c,i)) | c \in \mathcal{C}\}$$

Any labeled relation between any pair of concepts $(c, c') \in \mathcal{C} \times \mathcal{C}'$ is represented by a triple $((c,c'), l, \#(c,c',l,i))$, where l is an element of $\mathcal{L}$, the set of possible labels for the relation, and $\#(c,c',l,i)$ is the number of times c and c' are related with label l in image i. W^i_E is then defined as:

$$W^i_E = \{((c,c'), l, \#(c,c',l,i)) | (c,c') \in \mathcal{C} \times \mathcal{C}', l \in \mathcal{L}\}$$

If a pair of concepts (c, c') come from the same concept set, we refer this relation as *intra-relation* set. Otherwise, we refer it as *inter-relation* set.

1.2 Language model for graph matching

Inspired by [1], the probability for a graph query $G^q =< S^q_{WC}, S^q_{WE} >$ to be generated from one graph document G^d is computed as:

$$P(G^q|G^d) \quad = \quad P(S^q_{WC}|G^d) \times P(S^q_{WE}|S^q_{WC}, G^d) \quad (1)$$

This probability composed of two parts: a probability to generate the concept sets $P(S^q_{WC}|G^d)$ and a probability to generate the relation sets $P(S^q_{WE}|S^q_{WC}, G^d)$. The probability of generating query concept sets from the document model $P(S^q_{WC}|G^d)$ uses a concept set independence hypothesis:

$$P(S^q_{WC}|G^d) = \prod_{W^q_C \in S^q_{WC}} P(W^q_C|G^d) \quad (2)$$

Assuming concept independence, standard in information retrieval and the number of occurrences of the concepts (i.e., the weights considered previously) are integrated through the use of a multinomial model, we compute $P(W^q_C|G^d)$ as:

$$P(W^q_C|G^d) \quad \propto \quad \prod_{c \in \mathcal{C}} P(c|G^d)^{\#(c,q)} \quad (3)$$

where $\#(c, q)$ denotes the number of times concept c occurs in the graph representation of the query. This contribution corresponds to the concept probability. The quantity $P(c|G^d)$ can be estimated through maximum likelihood using Jelinek-Mercer smoothing:

$$P(c|G^d) = (1-\lambda_C)\frac{\#(c,d_C)}{\#(*,d_C)} + \lambda_C \frac{\#(c,D_C)}{\#(*,D_C)} \quad (4)$$

where $\#(c, d)$ represents the number of occurrences of c in the graph representation of the image d, and where $\#(*, d)$ is equal to $\sum_c \#(c, d)$. The quantities $\#(c, D)$ are similar, but defined over the whole collection (i.e., over the union of all images in the collection). Based on the relation set independence hypothesis, we follow a similar process for the relation sets, leading to:

$$P(S^q_{WE}|S^q_{WC}, G^d) = \prod_{W^q_E \in S^q_{WE}} P(W^q_E|S^q_{WC}, G^d) \quad (5)$$

For the probability of generating query relation from the document, we assume that a relation depends only on the two linked sets. Assuming that the relations are independent and following a multinomial model, we compute:

$$P(W^q_E|S^q_{WC}, G^d) \propto \prod_{(c,c',l)\in\mathcal{C}\times\mathcal{C}'\times\mathcal{L}} P(L(c,c') = l|W^q_C, W^q_{C'}, G^d)^{\#(c,c',l,q)} \quad (6)$$

where $c \in \mathcal{C}$, $c' \in \mathcal{C}'$ and $L(c, c')$ is a variable with values in $\mathcal{L}$ reflects the possible relation labels between c and c', in this relation set. The parameters of the model $P(L(c, c') = l|W^q_C, W^{q\prime}_C, G^d)$ are estimated by the maximum likelihood with Jelinek-Mercer smoothing. Images are ranked based on their relevance status value.

2. EXPERIMENTAL RESULTS

In order to assess the validity of our methods, we have experimented on 2 image collections and compared the results with other state-of-the-art approach in image categorization such as SVM classification method (implemented thanks to the *libsvm*[1]). We applied the same visual features used for our experiment. Each class was trained with a corresponding SVM classifier using RBF kernel.

STOIC-101 collection

The STOIC-101 collection contains 3849 photos of 101 tourist landmarks in Singapore. For experimental purposes, the collection has been divided into a training set of 3189 images and a test set of 660 images. We extracted from each block of 10×10 pixels a center pixel as a representative for the region. From this pixel, a vector of HSV color (8 bins for each channel) was extracted and clustered into 500 concepts. Based on this representation, we built 2 graph models: (1) with only concept set, referred as simple Language Model (LM); (2) with integrating of intra-relation set $\{left\_of, top\_of\}$ to concept set, referred as Visual Graph Model (VGM). As a comparison, we present in table 1 the best results obtained using SVM classifiers.

RobotVision'09 collection

The RobotVision'09 collection was used for ImageCLEF competition aiming to address the problem of localization of a robot using only the visual information. This collection contains a sequence of 1034 images for training and a sequence of 909 images for validation. The official test is carried out on a set of 1690 images. Image sequences were captured within an indoor laboratory environment consisting of 5 rooms. For this collection, we have applied 2 types of image representations: (1) regular division of 5×5 patches; (2) extraction of SIFT (Scale Invariant Feature Transform) features from local key-points. From these image representations, we defined an inter-relation set $\{inside\}$ between patches and key-points representation if one key-point is located **inside** the region of one patch. Similar to above, we referred the model without relation as LM (simply the production of probability generated by different concept sets) and graph model with the spatial relation as VGM (with the contributing of relation probability to graph model). The SVM model was trained based on the fusion of concepts come from both patches and SIFT key-point features.

[1]http://www.csie.ntu.edu.tw/cjlin/libsvm/

Table 1: Results on categorizing STOIC-101 and RobotVision'09 collections

	$\#class$	LM	**VGM**	SVM
STOIC-101	101	0.789	**0.809** (+2.5%)	0.744
RobotVision				
Validation	5	0.579	**0.675** (+16.6%)	0.535
Test	6	0.416	**0.449** (+7.9%)	0.439

Table 1 summarizes the results obtained from both collection STOIC-101 and RobotVision'09. We can see that in all cases our VGMs outperformed other methods. More precisely, with the integration of spatial relation into VGM helped improving the accuracy of classical approaches of LM by at least 2.5%. Especially with the RobotVision collection, VGMs have increased roughly the accuracies of 7.9% to 16.6% comparing to LMs respectively for both test and validation set. Lastly, the VGMs have retained medium to large improvements over the state-of-the-art SVM classifiers in both image collections.

3. CONCLUSION

In this work, we have presented a novel graph-based framework for integrating smoothly the spatial relationships of visual concepts. Our contributions are two-folds: (1) a well-foundeness graph model for representation of image content (2) a simpler and more effective graph matching process based on the language model. Our experimental results confirmed the stability of our visual graph models, as well as, enhanced the results obtained with other approaches such as standard LM and SVM classification method.

Acknowledgments

This work was supported by AVEIR (ANR-06-MDCA-002) and Merlion PhD. programme from Singapore.

4. REFERENCES

[1] L. Maisonnasse, E. Gaussier, and J. Chevalet. Model fusion in conceptual language modeling. In *ECIR'09*, pages 240–251, 2009.

[2] J. M. Ponte and W. B. Croft. A language modeling approach to information retrieval. In *SIGIR'98*, 1998.

[3] P. Tirilly, V. Claveau, and P. Gros. Language modeling for bag-of-visual words image categorization. In *CIVR'08*, pages 249–258, 2008.