Spatial Emerging Patterns for Scene Classification

Łukasz Kobyliński and Krzysztof Walczak

Institute of Computer Science, Warsaw University of Technology
ul. Nowowiejska 15/19, 00-665 Warszawa, Poland
{L.Kobylinski, K.Walczak}@ii.pw.edu.pl

Abstract. In this paper we propose a novel method of scene classification, based on the idea of mining emerging patterns between classes of images, represented in a symbolic manner. We use the 9DLT (Direction Lower Triangular) representation of images, which allows to describe scenes with a limited number of symbols, while still capturing spatial relationships between objects visible on the images. We show an efficient method of mining the proposed Spatial Emerging Patterns and present results of synthetic image classification experiments.

1 Introduction

The approaches to mining data in image databases vary greatly in the way spatial data is represented and used for reasoning about its content. Recently, the attention is aimed at extracting local features from images, predominantly using the SIFT descriptor [1] and its descendants. For example in [2] the authors represent the image as a collection of local regions, extracted using a feature descriptor and clustered into a chosen number of unique codewords. Another recent advancement in image mining consists of analyzing the spatial relationships between local features and taking into consideration the context of identified objects. In [3] the images are partitioned into several layers of sub-regions and histograms of local features inside the sub-regions are calculated. This extension of a bag-of-words model makes an assumption that similar parts of a scene always occur in the same parts of the image grid. In the approach chosen in [4] the object context is taken into consideration by using a conditional random-field framework and this way the categorization accuracy is greatly improved.

Independently from image understanding methods, a great number of data mining approaches for transactional databases have been developed in recent years. The association rules and, more recently, emerging patterns, are just two examples of fruitful research in the area of data analysis. As these methods proved to perform well in the area of market basket analysis, text analysis and mining of other symbolic data, a question arises whether they can be used in image understanding. Such an application requires that the images are represented symbolically, usually by first extracting their features and employing an unsupervised learning method to get a number of codewords used to describe the scene. The representation may be created at various concept levels, however, beginning with individual pixels, through low-level features to real-world objects. Some of the first applications of data mining methods to discovery of knowledge in image databases has been proposed in [5], where association rules are mined between objects appearing in an image. The spatial relationships between features and objects have also been incorporated into proposed mining methods. One approach is to include the information about spatial context into the symbolic representation itself. In the 9DLT representation [6] the relationships between objects are denoted by associating directional codes with pairs of items, which provide information about the angle between two image features.

In our work we employ the very promising idea of mining emerging patterns in an image database consisting of scenes with identified objects, described symbolically with the 9DLT representation. In such a framework we may reason about spatial arrangement of objects visible on an image accurately and efficiently and use this knowledge in classification of scenes. This is done by mining a new type of emerging patterns – jumping spatial emerging patterns – in a database of symbolically represented training images with known category labels and then using this knowledge to classify previously unknown scenes.

2 Previous Work

The idea of using jumping emerging patterns for classification and their mining algorithm has been first proposed in [7]. Efficient discovery of emerging patterns has been studied in [8], while in [9] an efficient algorithm of mining jumping emerging patterns with recurrent items has been proposed. Such patterns proved to be useful in classification of multimedia data.

A method of discovering spatial association rules in image data has been proposed in [10]. The authors have used the 9DLT symbolic representation of images to allow mining interesting association rules between objects appearing in a visual database.

3 Emerging Patterns

Emerging patterns may be briefly described as patterns, which occur frequently in one set of data and seldomly in another. We now give a formal definition of emerging patterns in transaction systems.

Let a transaction system be a pair $(\mathcal{D}, \mathcal{I})$, where $\mathcal{D}$ is a finite sequence of transactions $(T_1, \ldots, T_n)$ (database), such that $T_i \subseteq \mathcal{I}$ for $i = 1, \ldots, n$ and $\mathcal{I}$ is a non-empty set of items (itemspace). A support of an itemset $X \subset \mathcal{I}$ in a sequence $D = (T_i)_{i \in K \subseteq \{1,\ldots,n\}} \subseteq \mathcal{D}$ is defined as: $\text{supp}_D(X) = \frac{|\{i \in K\colon\ X \subseteq T_i\}|}{|K|}$.

Let a decision transaction system be a tuple $(\mathcal{D}, \mathcal{I}, \mathcal{I}_\mathrm{d})$, where $(\mathcal{D}, \mathcal{I} \cup \mathcal{I}_\mathrm{d})$ is a transaction system and $\forall_{T \in \mathcal{D}} |T \cap \mathcal{I}_\mathrm{d}| = 1$. Elements of $\mathcal{I}$ and $\mathcal{I}_\mathrm{d}$ are called condition and decision items, respectively. A support for a decision transaction system $(\mathcal{D}, \mathcal{I}, \mathcal{I}_\mathrm{d})$ is understood as a support in the transaction system $(\mathcal{D}, \mathcal{I} \cup \mathcal{I}_\mathrm{d})$.

For each decision item $c \in \mathcal{I}_\mathrm{d}$ we define a decision class sequence $C_c = (T_i)_{i \in K}$, where $K = \{k \in \{1, \ldots, n\} : c \in T_k\}$. Notice that each of the transactions from $\mathcal{D}$ belongs to exactly one class sequence. In addition, for a database $D = (T_i)_{i \in K \subseteq \{1,\ldots,n\}} \subseteq \mathcal{D}$, we define a complement database $D' = (T_i)_{i \in \{1,\ldots,n\} - K}$.

Given two databases $D_1, D_2 \subseteq \mathcal{D}$ the growth rate of an itemset $X \subset \mathcal{I}$ from D_1 to D_2 is defined as:

$$GR_{D_1 \to D_2}(X) = \begin{cases} 0 & \text{if } \mathrm{supp}_{D_1}(X) = 0 \text{ and } \mathrm{supp}_{D_2}(X) = 0, \\ \infty & \text{if } \mathrm{supp}_{D_1}(X) = 0 \text{ and } \mathrm{supp}_{D_2}(X) \neq 0, \\ \frac{supp_{D_2}(X)}{supp_{D_1}(X)} & \text{otherwise.} \end{cases} \qquad (1)$$

Given a minimum growth rate ρ, we define an itemset $X \subset \mathcal{I}$ to be a ρ-emerging pattern (ρ-EP) from D_1 to D_2 if $GR_{D_1 \to D_2}(X) > \rho$. Furthermore, we say that an itemset X is a jumping emerging pattern (JEP), when its growth rate is infinite, that is $GR_{D_1 \to D_2}(X) = \infty$. Having a minimum support threshold ξ, we define a ξ-strong jumping emerging pattern to be a JEP from D_1 to D_2 for which $\mathrm{supp}_{D_1}(X) = 0$ and $\mathrm{supp}_{D_2}(X) > \xi$. A set of all JEPs from D_1 to D_2 is called a JEP space and denoted by $JEP(D_1, D_2)$.

4 Image Representation

We use the 9DLT string representation of images to capture the spatial arrangement of objects visible in a scene. The symbolic representation, which consists of object labels and directional codes indicating spatial relationships between them, allows us to use data mining methods to reason about large image databases.

The 9DLT representation defines nine directional codes, $\mathcal{R} = \{0, 1, \ldots, 8\}$, which are an equivalent of a range of angles between two objects in a scene. Figure 1a depicts the use of codes: "0" means "the same spatial location as", "1" means "the north of", "2" means "the north-west of", and so on.

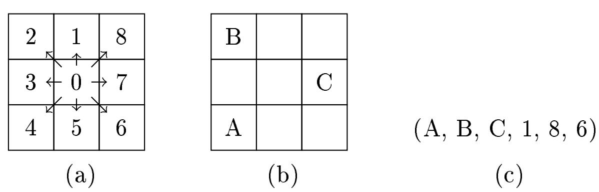

Fig. 1: The 9DLT representation: (a) directional codes, (b) example scene, (c) its symbolic representation.

We now use the definition of a spatial pattern, presented in [10], to extend the definition of our transactional system. A spatial pattern X^{s} is defined as a pattern of a form $X^{\mathrm{s}} = (i_1, i_2, \ldots, i_n, r_1, r_2, \ldots, r_m)$, where $i_j \in \mathcal{I}$ are items and $r_k \in \mathcal{R}$ are directional codes. Here, $m = C_2^n = n(n-1)/2)$, $1 \leq j \leq n$, $1 \leq k \leq m$ and $n \geq 2$. Each of the directional codes denotes a spatial relationship between two corresponding items, taken from left to right, e.g. the relationship between i_1 and i_2 is r_1, while between i_1 and i_3 is r_2.

Example 1. Consider the image presented on Fig. 1b. Its symbolic representation as a spatial pattern takes the form shown on Fig. 1c.

We say that spatial pattern $Y^{\mathrm{s}} = (i'_1, i'_2, \ldots, i'_n, r'_1, r'_2, \ldots, r'_m)$ is a sub-pattern of a pattern $X^{\mathrm{s}} = (i_1, i_2, \ldots, i_n, r_1, r_2, \ldots, r_m)$, denoted as $Y^{\mathrm{s}} \sqsubseteq X^{\mathrm{s}}$, when $\{i'_1, i'_2, \ldots, i'_n\} \subseteq \{i_1, i_2, \ldots, i_n\}$ and each spatial relationship between every two items is exactly the same in both patterns. Furthermore, we say that two spatial relationships $r_i, r_j \neq 0$ are complementary, when $r_i = (r_j + 4) \bmod 8$.

Example 2. Consider the following 4-element spatial pattern: $X^{\mathrm{s}} = (A, B, C, D, 8, 7, 8, 6, 1, 2)$. There are four 3-element sub-patterns of pattern X^s: $Y_1^s = (A, B, C, 8, 7, 6)$, $Y_2^s = (A, B, D, 8, 8, 1)$, $Y_3^s = (A, C, D, 7, 8, 2)$ and $Y_4^s = (B, C, D, 6, 1, 2)$.

A spatial transactional system is a pair $(\mathcal{D}^{\mathrm{s}}, \mathcal{I})$, where $\mathcal{D}^{\mathrm{s}}$ is a finite sequence of transactions $(T_1^{\mathrm{s}}, \ldots, T_n^{\mathrm{s}})$ for $i = 1, \ldots, n$. Each transaction is a pattern $(i_1, i_2, \ldots, i_n, r_1, r_2, \ldots, r_m)$, where $i_j \in \mathcal{I}$ are items and $r_k \in \mathcal{R}$ are directional codes. A support of a spatial pattern X^{s} in a sequence $D^{\mathrm{s}} = (T_i^{\mathrm{s}})_{i \in K \subseteq \{1,\ldots,n\}} \subseteq \mathcal{D}^{\mathrm{s}}$ is defined as:

$$\mathrm{supp}_{D^{\mathrm{s}}}(X) = \frac{|\{i \in K : \ X^{\mathrm{s}} \sqsubseteq T_i^{\mathrm{s}}\}|}{|K|}. \tag{2}$$

5 Spatial Emerging Patterns

5.1 Formal Definition

Based on the earlier definitions, we now define a new kind of patterns, namely Spatial Emerging Patterns (SEPs), which are able to capture interesting differences between sets of spatial data. Given two spatial databases D_1^{s} and D_2^{s}, we define the growth rate of a pattern X^{s} in the same way as stated by Eq. 1, in which we use the definition of support presented by Eq. 2. Having a minimum growth rate ρ, we define a pattern X^{s} to be a ρ-spatial emerging pattern (ρ-SEP) from D_1^{s} to D_2^{s} if $GR_{D_1^{\mathrm{s}} \to D_2^{\mathrm{s}}}(X) > \rho$. The definition of a jumping spatial emerging pattern (JSEP) and a ξ-strong jumping spatial emerging pattern is analogous to the one proposed for regular EPs.

We may introduce another way of representing spatial emerging patterns, which shows the connection between SEPs and regular emerging patterns. By enumerating all encoded relationships and creating unique item for each of them, we get a new space of items, which is defined as $\mathcal{I}' = \mathcal{I} \times \mathcal{R} \times \mathcal{I}$. Formally, we can say that each pattern of a form $X^{\mathrm{s}} = (i_1, i_2, \ldots, i_n, r_1, r_2, \ldots, r_m)$ may be represented as: $X^{\mathrm{s}} = (i_1 i_2 r_1, i_1 i_3 r_2, \ldots, i_1 i_n r_k, \ldots, i_{n-1} i_n r_m)$.

Example 3. A pattern $X^{\mathrm{s}} = (A, B, C, 1, 8, 6)$ may also be represented as $X^{\mathrm{s}} = (AB1, AC8, BC6)$, written for convenience as $X^{\mathrm{s}} = (A_1B, A_8C, B_6C)$.

It is important to note, that while all patterns may be represented in the second manner, not all patterns may be described in the original, shortened form. It is the case when not all relationships between particular items are known.

Example 4. Consider two sets of spatial data, represented by 9DLT patterns: $D_1 = ((A, B, C, 1, 8, 6)) = ((A_1B, A_8C, B_6C))$ and $D_2 = ((A, B, 1), (A, C, 8), (B, C, 7)) = ((A_1B), (A_8C), (B_7C))$. We mine strong JSEPs between these sets by looking for minimal patterns, which occur in one set and never in the other. In the case of JSEPs from D_1 to D_2 we have $JSEP_1 = (B, C, 6) = (B_6C)$ and $JSEP_2 = (A, B, C, 1, 8, ?) = (A_1B, A_8C)$. Similarly, in the direction of D_2 to D_1 we have $JSEP_3 = (B, C, 7) = (B_7C)$.

5.2 Mining Algorithm

In our approach, we are only interested in mining patterns for the use in building classifiers. For that reason we may limit ourselves to mining only strong jumping spatial emerging patterns, that is JSEPs, which are minimal and have a specified minimum support in one of the databases.

An efficient algorithm for mining emerging patterns has been presented in [8], which introduces the notion of borders to represent a large number of patterns. A border is an ordered pair $< \mathcal{L}, \mathcal{R} >$ such that $\mathcal{L}$ and $\mathcal{R}$ are antichains, $\forall X^{\mathrm{s}} \in \mathcal{L}\ \exists Y^{\mathrm{s}} \in \mathcal{R}\ X^{\mathrm{s}} \subseteq Y^{\mathrm{s}}$ and $\forall X^{\mathrm{s}} \in \mathcal{R}\ \exists Y^{\mathrm{s}} \in \mathcal{L}\ Y^{\mathrm{s}} \subseteq X^{\mathrm{s}}$. The collection of sets represented by a border $< \mathcal{L}, \mathcal{R} >$ is equal to:

$$[\mathcal{L}, \mathcal{R}] = \{Y^{\mathrm{s}} : \exists X^{\mathrm{s}} \in \mathcal{L}, \exists Z^{\mathrm{s}} \in \mathcal{R} \text{ such that } X^{\mathrm{s}} \subseteq Y^{\mathrm{s}} \subseteq Z^{\mathrm{s}}\}. \tag{3}$$

The left border in this representation corresponds to minimal patterns found in a particular dataset. As such, we may follow the methodology of finding only the left border of the set of jumping spatial emerging patterns between the two databases. Having databases D_1 and D_2 this may be performed by subtracting all patterns in D_2 from each of the patterns in D_1 and vice versa. By aggregating all the resulting patterns, we get the set of minimal JSEPs from D_2 to D_1 and from D_1 to D_2 respectively.

The straightforward way of calculating this differential would be to find all sub-patterns of each of the database transactions and eliminate these patterns in R_1, which also occur in R_2. To avoid the cost of checking all possible relationships in both databases, an iterative procedure may be used, which comes from the idea presented in [8]. It has been shown there that the collection of minimal itemsets $\mathrm{Min}(\mathcal{S})$ in a border differential $\mathcal{S} = [\{\emptyset\}, \{U^{\mathrm{s}}\}] - [\{\emptyset\}, \{S^{\mathrm{s}}_1, \ldots, S^{\mathrm{s}}_k\}]$ is equivalent to: $\mathrm{Min}(\{\bigcup\{s_1, \ldots, s_k\} : s_i \in U^{\mathrm{s}} - S^{\mathrm{s}}_i, 1 \le i \le k\})$.

We may iteratively expand candidate patterns and check if they are minimal, avoiding in this way generating many unnecessary patterns. The complete procedure is presented as Algorithm 1 below. We need to iteratively call the Border-differential function and create a union of the results to find the set of all minimal jumping spatial emerging patterns from C'_c to C_c.

5.3 Scene Classification using Spatial Emerging Patterns

Having discovered spatial patterns on the basis of the learning set, we may use the built classifier to categorize previously unseen images from the testing set. To

Algorithm 1: Border differential

Input : $<\{\emptyset\},\{U^{\mathrm{s}}\}>, <\{\emptyset\},\{S_1^{\mathrm{s}},\ldots,S_k^{\mathrm{s}}\}>$
Output: $\mathcal{L}$

$T_i^{\mathrm{s}} \longleftarrow U^{\mathrm{s}} - S_i^{\mathrm{s}}$ **for** $1 \le i \le k$
if $\exists T_i^{\mathrm{s}} = \{\emptyset\}$ **then**
  return $<\{\},\{\}>$
end
$\mathcal{L} \longleftarrow \{\{x\} : x \in T_1^{\mathrm{s}}\}$
for $i = 2$ ***to*** k **do**
  $NewL \longleftarrow \{X^{\mathrm{s}} \in \mathcal{L} : X^{\mathrm{s}} \cap T_i^{\mathrm{s}} \neq \emptyset\}$
  $\mathcal{L} \longleftarrow \mathcal{L} - NewL$
  $T_i^{\mathrm{s}} \longleftarrow T_i^{\mathrm{s}} - \{x : \{x\} \in NewL\}$
  foreach $X^{\mathrm{s}} \in \mathcal{L}$ *sorted according to increasing cardinality* **do**
    foreach $x \in T_i$ **do**
      if $\forall Z^{\mathrm{s}} \in NewL\ \mathrm{supp}_{Z^{\mathrm{s}}}(X^{\mathrm{s}} \cup \{x\}) = 0$ **then**
        $NewL \longleftarrow NewL \cup (X^{\mathrm{s}} \cup \{x\})$
      end
    end
  end
  $\mathcal{L} \longleftarrow NewL$
end

classify a particular scene, we first transform the image into its symbolic form, using the 9DLT representation. Next, we aggregate all minimal JSEPs, which are supported by the representation. A scoring function is calculated and a category label is chosen by finding the class with the maximum score: $\mathrm{score}(T^{\mathrm{s}}, c) = \sum_{X^{\mathrm{s}}} \mathrm{supp}_{C_c}(X^{\mathrm{s}})$, where $C_c \subseteq \mathcal{D}_{\mathrm{T}}^{\mathrm{s}}$ and $X^{\mathrm{s}} \in JSEP_{\mathrm{m}}(C'_c, C_c)$, such that $X^{\mathrm{s}} \subseteq T^{\mathrm{s}}$.

6 Experimental Results

To assess the effectiveness of the proposed method, we have performed experiments using synthetic data to build and then test the JSEP-based classifiers. The data is prepared as follows: we generate two classes of transactions, consisting of uniformly distributed objects in a $n \times n$ image. For each of the classes a characteristic pattern of size $m \times m$, $m < n$ is randomly constructed and overlaid on each of the random images. Finally, the images are transformed to 9DLT strings. This way we can assess the performance of the described classification method in recognizing the differentiating pattern in a set of otherwise random data. Apart from the image and pattern sizes the following parameters of the data generator may be changed freely: number of available objects (K), number of objects on a single image (L) and number of transactions in each of the classes (D). Having generated the synthetic dataset we perform a ten-fold cross-validation experiment, by first discovering the jumping spatial emerging patterns in 90% of available data and testing the accuracy of classification in the other 10%. This procedure is repeated 10 times and an average accuracy is presented in results below.

Firstly, we have experimented with the influence of the relation between pattern and image sizes on classification accuracy and the time needed to mine spatial patterns. The results are presented in Table 1 and show an increase of accuracy when pattern size approaches the size of the image. This is because there is relatively less random noise in the generated data in comparison to the differentiating pattern. The image size alone however, does not directly influence the classification accuracy or pattern mining time, as it has no relation to the size of 9DLT representation and number of discovered JSEPs.

Table 1: Classification accuracy of the synthetic dataset with relation to image and pattern sizes.

Image size (n)	Accuracy (%)	Time (ms)
4	95,50	1738
5	92,20	2790
6	94,30	3218
7	92,70	3607
8	95,00	3752
9	93,10	3934
10	92,90	3653
$K = 10$, $L = 5$, $D = 100$, $m = 3$		

Pattern size (m)	Accuracy (%)	Time (ms)
2	82,00	2397
3	95,00	3545
4	98,00	5840
5	98,50	8109
$K = 10$, $L = 5$, $D = 100$, $n = 8$		

The influence of the size of object space and the number of objects appearing on a particular image on classification results may be assessed from the data in Table 2. We can see that increasing the object space size while maintaining a constant number of objects results in better classification accuracy and less time needed to mine JSEPs, while increasing the number of objects having a constant objects space size has an opposite effect. The number of objects on individual images has a direct impact on the average length of the 9DLT representation and thus the transaction size. As the average transaction length increases, the number of patterns becomes larger and so does the ratio between noise and the patterns which allow to differentiate between classes.

7 Conclusions and Future Work

In this paper we have introduced Spatial Emerging Patterns, a new data mining method of discovering knowledge in image databases and its use in classification of scenes. The presented results of the experiments look promising and show that the method may be used to classify image data, in which a preliminary object recognition step has been performed. Such images may be transformed into 9DLT representation and used as a basis for JSEP mining and classification.

Table 2: Classification accuracy of the synthetic dataset with relation to the number of available objects and objects on a single image.

Object space (K)	Accuracy (%)	Time (ms)
10	93,00	3663
15	97,80	2604
20	98,30	1637
25	98,50	1264
30	99,50	935
40	99,83	923
50	100,00	911
$L = 5$, $D = 100$, $n = 8$, $m = 3$		

Number of objects (L)	Accuracy (%)	Time (ms)
3	96,70	376
4	96,00	1549
5	92,40	3663
6	86,00	9116
7	81,20	18968
8	78,50	42730
$K = 10$, $D = 100$, $n = 8$, $m = 3$		

Other symbolic representations may be proposed in the place of 9DLT and it remains a further work to assess the influence of the representation methods used on overall accuracy.

References

1. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60**(2) (2004) 91–110
2. Li, F.F., Perona, P.: A bayesian hierarchical model for learning natural scene categories. In: Proc. 2005 IEEE Conf. Computer Vision and Pattern Recognition (CVPR'05), Washington, DC, USA, IEEE Computer Society (2005) 524–531
3. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: Proc. 2006 IEEE Conf. Computer Vision and Pattern Recognition (CVPR'06), Washington, DC, USA, IEEE Computer Society (2006) 2169–2178
4. Rabinovich, A., Vedaldi, A., Galleguillos, C., Wiewiora, E., Belongie, S.: Objects in context. In: 11th Int. Conf. Comp. Vision (ICCV'07), Rio de Janeiro (2007) 1–8
5. Ordonez, C., Omiecinski, E.: Discovering association rules based on image content. In: Proc. IEEE Forum on Research and Technology Advances in Digital Libraries (ADL'99), Washington, DC, USA, IEEE Computer Society (1999)
6. Chan, Y., Chang, C.: Spatial similarity retrieval in video databases. Journal of Visual Communication and Image Representation **12** (2001) 107–122
7. Li, J., Dong, G., Ramamohanarao, K.: Making use of the most expressive jumping emerging patterns for classification. Knowledge and Information Systems **3**(2) (2001) 1–29
8. Dong, G., Li, J.: Mining border descriptions of emerging patterns from dataset pairs. Knowledge and Information Systems **8**(2) (2005) 178–202
9. Kobyliński, Ł., Walczak, K.: Efficient mining of jumping emerging patterns with occurrence counts for classification. In Chan, C.C., Grzymała-Busse, J.W., Ziarko, W.P., eds.: Int. Conf. Rough Sets and Current Trends in Computing (RSCTC'08). Volume 5306 of LNAI., Springer-Verlag (2008) 419–428
10. Lee, A.J.T., Hong, R.W., Ko, W.M., Tsao, W.K., Lin, H.H.: Mining spatial association rules in image databases. Information Sciences **177**(7) (2007) 1593–1608