

# Extended Symbolic Mining of Textures with Association Rules

Igor Kononenko and Matjaž Bevk

Faculty of computer and information science, University of Ljubljana, Tržaška 25

E-mail: igor.kononenko@fri.uni-lj.si and lkm.fri.uni-lj.si/xaigor

**Keywords:** texture description, association rules, image mining, multi-resolution, meta-association rules, ArTex algorithm

**Received:** January 25, 2008

*The association rules algorithms can be used for describing textures if an appropriate texture representation formalism is used. This representation has several good properties like invariance to global lightness and invariance to rotation. Association rules capture structural and statistical information and are very convenient to identify the structures that occur most frequently and have the most discriminative power. This paper presents the extended textural features which are based on association rules. We extend the basic textural features of our algorithm ArTex in three ways, using (a) various interestingness measures, (b) the multi-resolution approach, and (c) the meta-parameters. Results of our experiments show that the extended representation improves the utility of basic textural features and often gives better results, comparable to standard texture descriptions.*

*Povzetek: Z metodami asociacijskih pravil je razvit algoritem za analizo tekstur.*

## 1 Introduction

Texture is a commonly used term in computer vision, although it is difficult to precisely define it. We can regard an image texture as a function of spatial variation in pixel values. There exist many different approaches to characterize textures. Most texture features are based on structural, statistical or spectral properties of the image. Well known statistical features are based on gray-level cooccurrence statistics[12]. Examples of structural features are features of Voronoi tessellation[27], representations using graphs[28], representations using grammars[22] and representations using association rules[23]. We have developed[2] a similar approach to that of Rushing et al.[23]. We showed that our approach, implemented with algorithm ArTex, as opposed to that of Rushing et al., is rotation-invariant as well as brightness-invariant, produces significantly smaller descriptions and is significantly faster[2]. Due to excellent efficiency, we decided to upgrade our approach with additional descriptors that may lead to better descriptions of target images, although with an increased time and description complexity. This paper presents the extended textural features, based on association rules, of algorithm ArTex, as described in Ref. [2]. We extend the basic textural features in three ways, using (a) various interestingness measures, (b) the multi-resolution approach, and (c) the meta-parameters. We show that the extended parameters together with the basic ones are appropriate for effective description of images and can be efficiently induced. The purpose of using association rules is to obtain a structural description of textures. The ultimate goal, which we hope to reach in further work, is to get higher order association rules, which would capture the

global structure of the image and would also allow for a transparent (human readable) description of the image. The present study is a step towards that goal.

This paper is organized as follows: Section 2 gives a brief definition of association rules, describes a texture representation which is suitable for processing with association rule algorithms, and shows how association rules are used for feature description of textures with algorithm ArTex. In Section 3 we define the three extensions to the basic feature set. The following section shows a comparison between the extended feature set and the basic one, as well as the comparison with other algorithms for image description. In Section 5 we show the results of the Extended ArTex in four real world data sets, and finally the last section concludes and gives ideas for further work.

## 2 Texture descriptions with association rules

### 2.1 Association rules

Association rules were introduced by Agrawal et al.[1] back in 1993. The following is a formal statement of the problem: Let  $\mathcal{I} = \{i_1, i_2, \dots, i_m\}$  be a set of literals, called items (also called transaction elements). Let  $\mathcal{D}$  be a set of transactions, where each transaction  $T$  is a set of transaction elements such that  $T \subseteq \mathcal{I}$ . We say that a transaction  $T$  contains  $B$ , if  $B \subseteq T$ . An *association rule* is an implication of the form  $B \implies C$ , where  $B \subset \mathcal{I}$ ,  $C \subset \mathcal{I}$  and  $B \cap C = \emptyset$ . The rule  $B \implies C$  holds in the transaction set  $\mathcal{D}$  with *confidence*  $c$  if  $c\%$  of transactions in  $\mathcal{D}$  that contain  $B$  also contain  $C$ . The rule  $B \implies C$  has *support*  $s$  in the

transaction set  $\mathcal{D}$  if  $s\%$  of transactions in  $\mathcal{D}$  contain  $B \cup C$ . The problem is to find all association rules in transaction set  $\mathcal{D}$  with confidence of at least *minconf* and support of at least *minsup*, where *minconf* and *minsup* represent the lower boundaries for confidence and support of association rules, respectively.

### 2.2 Texture representation

To apply association rules algorithms on textures one must first define the terms which are used in association rules in the context of textures.

**Pixel  $\vec{A}$  of a texture  $p$**  is a vector  $\vec{A} = (X, Y, I) \in p$ , where  $X$  and  $Y$  represent absolute coordinates and  $I$  represents the intensity of pixel  $A$ .

**Root pixel  $\vec{K}$**  is the current pixel of a texture  $\vec{K} = (X_K, Y_K, I_K)$ .

**$e$  neighborhood  $N_{e, \vec{K}}$**  is a set of pixels which are located in the circular area of radius  $e$  with root pixel  $\vec{K}$  at the center. Root pixel  $\vec{K}$  itself is not a member of its neighborhood.

$$N_{e, \vec{K}} = \{(X, Y, I) \mid \delta \leq e\} \setminus \vec{K}$$

$$\delta = \left\lceil \sqrt{(X_K - X)^2 + (Y_K - Y)^2} + 0.5 \right\rceil \quad (1)$$

**Transaction  $T_{e, \vec{K}}$**  is a set of elements based on its corresponding neighborhood. The elements of transaction are represented with Euclidean distance and the intensity difference from the root pixel.

$$T_{e, \vec{K}} = \left\{ (\delta, I_K - I) \mid (X, Y, I) \in N_{e, \vec{K}} \right\} \quad (2)$$

**Transaction element** is a two dimensional vector  $(r, i) \in T_{e, \vec{K}}$ , where the first component represents Euclidean distance from the root pixel and the second component represents the intensity difference from the root pixel.

**Association rule** is composed of transaction elements; therefore it looks like this

$$(r_1, i_1) \wedge \dots \wedge (r_m, i_m) \\ \implies (r_{m+1}, i_{m+1}) \wedge \dots \wedge (r_{m+n}, i_{m+n})$$

**Transaction set  $D_{p,e}$**  is composed of transactions, which are derived from all possible root pixels of a texture  $p$  at certain neighborhood size  $e$ . Possible root pixels are only those for which an  $e$  neighborhood would not extend outside of the texture’s borders.

$$D_{p,e} = \left\{ T_{e, \vec{K}} \mid \forall \vec{K} : \vec{K} \in p \right\}$$

This representation of a texture replaces exact information of the location and intensity of the neighboring pixels with

more vague information of distance and the relative intensity of neighboring pixels. This description is also rotation invariant.

The described representation is almost suitable for processing with general association rule algorithms. What is still to be considered is the form of a transaction element. The association rule algorithms expect scalar values for transaction elements, whereas our representation produces a two dimensional vector for a transaction element. Let us say that intensity of each texture point can have values from interval  $[0..(Q - 1)]$  and that the neighborhood size is  $e$ . Take some transaction element  $(r, i)$ , where  $i$  holds a value from  $[-(Q - 1)..+(Q - 1)]$  and  $r$  holds a value from  $[1..e]$ . What is needed here is a bijective mapping that transforms each vector to its scalar representation. A possible and quite straightforward solution is:

$$s = (2Q - 1)(r - 1) + i + (Q - 1) \quad (3)$$

Equation (3) maps each  $(r, i)$  pair into its unique scalar representation  $s$ , so that values from  $[0..2Q - 2]$  represent pixels which are located on distance 1 from the root pixel, values from  $[2Q - 1..4Q - 3]$  represent pixels which are located in distance 2 from the root pixel, and so on.

The transformation is also reversible:

$$r = 1 + s \operatorname{div} (2Q - 1) \\ i = s \operatorname{mod} (2Q - 1) - (Q - 1)$$

Now it is possible to define a transaction that suits the general association rule algorithms:

$$T_{e,Q, \vec{K}} = \left\{ s \mid \begin{array}{l} (r, i) \in T_{e, \vec{K}}, \\ s = (2Q - 1)(r - 1) + i + (Q - 1) \end{array} \right\}$$

And finally we obtain the appropriate transaction set definition:

$$\mathcal{D}_{p,e,Q} = \left\{ T_{e,Q, \vec{K}} \mid \forall \vec{K} : \vec{K} \in p \right\}$$

### 2.3 From association rules to feature description

We are ready to describe the algorithm ArTex[2]. Preprocessing consists of the following steps: conversion to gray scale, pixel quantization to  $Q$  levels (to make it faster and more accurate), and selection of the neighborhood size  $e$  (there is a trade-off between better descriptions with greater  $e$  and better time complexity with smaller  $e$ ).

It is important to understand that ArTex does not perform the final concept induction, it extracts features from images. These features are later used by external machine learning algorithm to induce the model (for classification of images). Let  $P$  represent the whole set of images. We isolate a small subset of images  $P_f \subset P$  which will be used for feature extraction, the rest of images  $P_l = P \setminus P_f$  will be used for learning. The selection of images for  $P_f$  is random, but it has to be ensured that each class is represented with equal amount of images.

### 2.3.1 Feature extraction with ArTex

The whole algorithm ArTex is given in Algorithm 1. The most important is the part that follows the preprocessing phase and the image subset selection phase. This part of the algorithm is denoted with lines from 6 to 12. Each image from  $p \in P_f$  is transformed into a transaction set  $\mathcal{D}_{p,e,Q}$ . The process of this transformation is described in Section 2.2. Transaction sets are then processed by algorithms AprioriN and GenRulesN, which return most important cooccurrences of pixels in the form of association rules.

### 2.3.2 Calculation of feature values

ArTex (Algorithm 1) is composed of two major parts. The first part (lines from 2 to 12) is the feature extraction, and the second part (lines from 13 to 16) is a process of evaluating feature values. The second part iterates through images from  $P_l$ , where features for each image are evaluated. There are two types of features: support and confidence for extracted association rules. We presume that we have pre-defined functions  $\rho_i(p, q)$  which evaluate feature  $q$  of type  $i$  (support or confidence) on image  $p$ . The result is stored in matrix  $d$ , where element  $d_{ij}$  represents the value of feature  $j$  on image  $i$ .

### 2.3.3 Modifications of Apriori and GenRules

Original algorithm Apriori[1] (line 8 in Algorithm 1) takes as a parameter the minimal required support (minsup) of transaction tuples. It is impossible to know in advance what is the appropriate value of minsup. If it is too large, no tuples will be found, and if it is too small, too many tuples will be generated which may drastically increase the time complexity (and also the description complexity) of ArTex. For that reason we use a slightly modified algorithm Apriori which as the input takes the number of tuples (ntup) to return, rather than the minimal support. It executes standard Apriori in an iterative manner, reducing the minimal support in each iteration. The modified Apriori is much more stable than standard Apriori and has no problems with time and description complexity.

Similarly as we modified Apriori, we also modified the GenRules algorithm, which is used to generate association rules from tuples[1] (line 9 in Algorithm 1). The same problem, as with minsup parameter in Apriori, algorithm Genrules has with the parameter minconf (minimal required confidence). We therefore use a modified GenRules algorithm which, instead of the minimal confidence, takes the minimal number of rules (minrul) to return.

## 3 Extending the default feature set

Our model of texture is such that the structure of association rules also describes some aspects of the textural structure. Since we are interested in parametric description of

```

1: procedure ARTEx( $P$  set of images,  $e$ ,  $Q$ , ntup, min-
   rul)
2:   for all  $p \in P$  do
3:     quantize pixels of image  $p$  to  $Q$  levels;
4:   end for
5:   select feature extraction subset  $P_f$  and learn/test
   subset  $P_l$ ,  $P = P_f \cup P_l$ ;
6:   for all  $p \in P_f$  do ▷ feature extraction
7:     represent  $p$  in transaction form  $\mathcal{D}_{p,e,Q}$ ;
8:      $r_{sup} = \text{apriori}(\mathcal{D}_{p,e,Q}, \text{ntup})$ ;
9:      $r_{conf} = \text{genrules}(r_{sup}, \text{minrul})$ ;
10:     $\rho_{sup} = \rho_{sup} \cup r_{sup}$ ;
11:     $\rho_{conf} = \rho_{conf} \cup r_{conf}$ ;
12:   end for
13:    $i = 0$ ;
14:   for all  $p \in P_l$  do ▷ calculate feature values for  $P_l$ 
15:      $j = 0$ ;
16:     for all  $\varrho \in \rho_{sup}$  do
17:        $d_{i,j} = \varphi_{sup}(p, \varrho)$ 
18:        $j = j + 1$ 
19:     end for
20:     for all  $\varrho \in \rho_{conf}$  do
21:        $d_{i,j} = \varphi_{conf}(p, \varrho)$ 
22:        $j = j + 1$ 
23:     end for
24:      $i = i + 1$ ;
25:   end for
26:   return  $d$ ; ▷ returns a matrix of extracted features  $d$ 
27: end procedure

```

**Algorithm 1:** Algorithm ArTex for feature extraction. Note that the number of features is determined dynamically and therefore the upper bounds for indices  $i$  and  $j$  cannot be determined in advance.

a texture, this structure has to be represented with one or more numerical parameters (numerical features).

### 3.1 Using various interestingness measures

Until now we presented the algorithm which uses only the basic interestingness measures support and confidence, which were defined together with association rules[1]. Nowadays, they are still most widely used, but studies have shown that there are some concerns especially with confidence measure, which can be misleading in many practical situations, as shown by Brin et al.[3]. They also offered an alternative to evaluate association rules using the  $\chi^2$  test. Contrary to confidence measure, the  $\chi^2$  test could be used to find both positively and negatively correlated association patterns. However, the  $\chi^2$  test alone may not be the ultimate solution because it does not indicate the strength of correlation between items of the association pattern. It only decides whether items of the association pattern are independent of each other, thus it cannot be used for ranking purposes.

We use  $\chi^2$  test just to select interesting association patterns, which are later described by the Pearson's correlation coefficient ( $\phi$ -coefficient) as advised in Ref. [25]. Besides  $\phi$ -coefficient, ArTex can also compute seven additional association rule interestingness measures, which are a subset of collection made by Tan et al.[26]. Table 1 lists interestingness measures that form an extended parameter set of the Extended ArTex.

Table 1: Extended set of interestingness measures for association rules ( $A \Rightarrow B$ ).  $P(X)$  represents the probability of  $X$ .

#	measure	formula
1	$\phi$ -coefficient	$\frac{P(A, B) - P(A)P(B)}{\sqrt{P(A)P(B)(1-P(A))(1-P(B))}}$
2	odds ratio	$\frac{P(A, B)P(\bar{A}, \bar{B})}{P(A, \bar{B})P(\bar{A}, B)}$
3	conviction	$\frac{P(A)P(\bar{B})}{P(A, \bar{B})}$
4	interest	$\frac{P(A, B)}{P(A)P(B)}$
5	Jaccard	$\frac{P(A, B)}{P(A) + P(B) - P(A, B)}$
6	Pia-Shapiro	$P(A, B) - P(A)P(B)$
7	J-measure	$P(A, B) \log\left(\frac{P(B A)}{P(B)}\right) + P(A, \bar{B}) \log\left(\frac{P(\bar{B} A)}{P(\bar{B})}\right)$
8	gini index	$\frac{P(A)[P(B A)^2 + P(\bar{B} A)^2] + P(\bar{A})[P(B \bar{A})^2 + P(\bar{B} \bar{A})^2] - P(B)^2 - P(\bar{B})^2}{2}$

It is possible now to define an upgraded version of ArTex - ArTexG - which extracts features based on an arbitrary set of interestingness measures. Besides other input parameters, ArTexG takes a list of interestingness measures  $\vec{\varphi}$  and their corresponding constraints  $\vec{t}$ . In the heart of the algorithm is  $Genrules_k$  function, which extracts best

rules according to  $k$ -th function from Table 1. This function is obtained by replacing the confidence criteria in the common *GenRules* algorithm with corresponding criteria from Table 1. When the first part of the algorithm finishes (lines from 1 to 14) it returns best association rules in sets  $\rho_{\varphi_k}$ , which were selected according to the selection criteria  $\varphi_k$ . The second part (lines from 15 to 31) is the feature evaluation process, which is very similar to original ArTex (Algorithm 1), except that it calculates feature values for all types of features  $\vec{\varphi}$ .

```

1: procedure ARTEXG( $P$  set of images,  $e, Q, \vec{\varphi}$  measures,
    $\vec{t}$  constraints)
2:   for all  $p \in P$  do
3:     quantize pixels of image  $p$  to  $Q$  levels;
4:   end for
5:   select feature extraction subset  $P_f$  and learn/test
   subset  $P_l$ ,  $P = P_f \cup P_l$ ;
6:   for all  $p \in P_f$  do  $\triangleright$  feature extraction
7:     represent  $p$  in transaction form  $\mathcal{D}_{p,e,Q}$ ;
8:      $r_{sup} = \text{apriori}(\mathcal{D}_{p,e,Q}, t_1)$ ;
9:      $\rho_{sup} = \rho_{sup} \cup r_{sup}$ ;
10:    for all  $\varphi_k$  do
11:       $r_{\varphi_k} = \text{genrules}_k(r_{sup}, t_k)$ ;  $\triangleright$  uses the
    $k$ -th interestingness measure
12:       $\rho_{\varphi_k} = \rho_{\varphi_k} \cup r_{\varphi_k}$ ;
13:    end for
14:  end for
15:   $i = 0$ ;
16:  for all  $p \in P_l$  do  $\triangleright$  calculate feature values for  $P_l$ 
17:     $j = 0$ ;
18:    for all  $\varrho \in \rho_{sup}$  do
19:       $d_{i,j} = \varphi_{sup}(p, \varrho)$ 
20:       $j = j + 1$ 
21:    end for
22:    for all  $\varphi_k$  do
23:      for all  $\varrho \in \rho_{\varphi_k}$  do
24:         $d_{i,j} = \varphi_k(p, \varrho)$ 
25:         $j = j + 1$ 
26:      end for
27:    end for
28:     $i = i + 1$ ;
29:  end for
30:  return  $d$ ;  $\triangleright$  returns a matrix of extracted features  $d$ 
31: end procedure

```

**Algorithm 2:** General algorithm ArTexG for feature extraction. Note that the number of features is determined dynamically and therefore the upper bounds for indices  $i$  and  $j$  cannot be determined in advance.

### 3.2 Using more than one resolution

Each induced association rule discovers a certain pattern in the textures. The next extension of parameters comes from the issue of the pattern's scale. Not every combination of scale and neighborhood size can guarantee that the pattern

would be detected. The problem is illustrated in Figures 1 and 2, where a fixed neighborhood  $e$  requires an appropriate resolution.

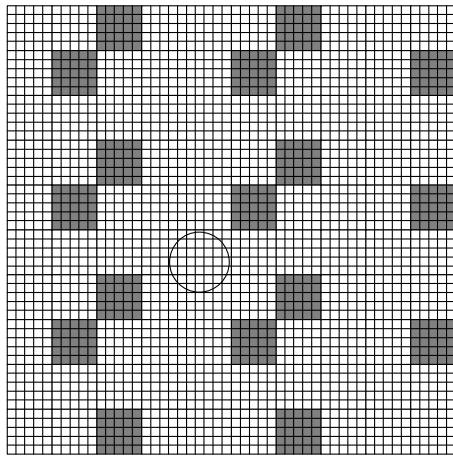


Figure 1: A very inadequate resolution for a fixed neighborhood.

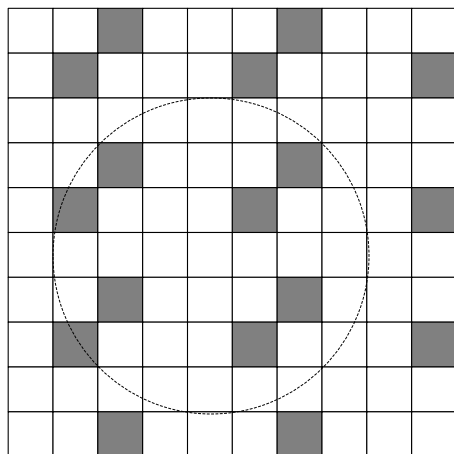


Figure 2: An adequate resolution for the example from Fig. 1.

We use the following definition: An image  $T_2$  of size  $M_2 \times N_2$  has a resolution  $\frac{1}{2}R$ , if and only if an image  $T_1$  of size  $M_1 \times N_1$  has resolution  $R$  where  $\frac{M_2}{M_1} = \frac{N_2}{N_1} = \frac{1}{2}$  and images  $T_1$  and  $T_2$  represent the same pattern.

To increase the possibility that the pattern will be detected, we propose a framework, where the extraction of association rules is repeated at different texture resolutions.

### 3.3 Meta parameters

To supply the learning algorithm with a more general view of the texture, we extended the parameter set with parameters based on meta association rules.

Let us split the learning set of images as follows:

$$P_l = P_{l_1} \cup P_{l_2},$$

Previously we defined  $P_f$  as a small subset that we use for feature extraction (see Section 2.3), thus set  $P_f \cup P_l$  represents all textures. Meta association rules are calculated on a learning subset of images  $P_{l_1}$ . Further, let us denote with  $\varphi_i$  the  $i$ -th interestingness measure, with  $\rho_{p_j, \varphi_i}$  an association rule that was discovered by using measure  $\varphi_i$  on texture  $p_j$ , and similarly with  $\rho_{*, \varphi_i}$  a set of association rules that was discovered by using measure  $\varphi_i$  on the feature extraction texture set  $P_f$ . The generation of meta-features is divided into three steps as follows:

1. First a set of binary features is constructed by using the default association rule set as follows. For each texture  $p \in P_{l_1}$  and for each interestingness measure  $\varphi_i$  and for each association rule  $\rho \in \rho_{*, \varphi_i}$  the value of the binary feature is 1 if association rule  $\rho$  was selected as important on texture  $p$  according to measure  $\varphi_i$ , otherwise the value of binary feature is 0.
2. This feature set along with the texture classes are then fed to an algorithm for generating associations rules, which induces rules with only the class on the consequent side. In our implementation we use algorithm Tertius[8]. Therefore, the output of Tertius is a set of rules of the form  $A \Rightarrow B$  where consequent  $B$  always consists only of the class, and features in the antecedent  $A$  are directly responsible for the outcome of the class. This way the output could easily be used for classification. Note that the “useful knowledge” of such a rule lies in the antecedent  $A$  which is a conjunctive combination of the binary features that were derived from the basic association rules.
3. The meta feature set is constructed by calculating the rules’ antecedents  $A$  on each texture from  $P_l$ . Note that the learning algorithm would have a trivial task and a strong bias if the learning set would consist of only of  $P_{l_1}$ , however, by adding  $P_{l_2}$  the task is hard enough. The introduced bias (hopefully) expresses the meta-regularity, discovered by the Tertius learning algorithm. Also note that meta features are binary features.

## 4 Experimental evaluation of the extended feature set

We performed experiments for comparing various versions of our algorithm with each other and with other texture description algorithms. For that purpose we used publicly available data bases.

### 4.1 Benchmark problem Domains

Here is a description of data bases used in our experiments:

- Outex[20]  
This data base contains a large variety of surface textures. The collection is well defined in terms of vari-



ations in illumination, rotation and spatial resolution. We chose the following collections of textures from the classification group:

- Outex 0  
The collection contains 480 images of 24 textures (classes). Each texture has 20 images. The dimension of images is  $128 \times 128$  pixels. The images contain no intentionally induced variation of illumination, rotation or spatial resolution.
- Outex 1  
The collection is similar to Outex 0, except that this collection contains 2112 images (88 per texture) and that they are of size  $64 \times 64$  pixels.
- Outex 2  
The collection is similar to Outex 0, except that this collection contains 8832 images (368 per texture) and that they are of size  $32 \times 32$  pixels.
- Outex 10  
The collection contains 4320 images, 24 textures; 180 images per texture. Textures are captured at different rotations of the surface:  $0^\circ, 5^\circ, 10^\circ, 15^\circ, 30^\circ, 45^\circ, 60^\circ, 75^\circ, 90^\circ$ . The size of images is  $128 \times 128$  pixels.
- Outex 11  
The collection contains 960 images, 24 textures; 40 images per texture. Textures are captured at two resolutions 100dpi and 120dpi. The size of images is  $128 \times 128$  pixels.
- Outex 12  
The collection contains 4800 images, 24 textures; 200 images per texture. The textures contain illumination variations and are captured at different rotations of the surface:  $0^\circ, 5^\circ, 10^\circ, 15^\circ, 30^\circ, 45^\circ, 60^\circ, 75^\circ, 90^\circ$ . The size of images is  $128 \times 128$  pixels.
- JerryWu[33]  
This data base was composed by Jerry Wu for his PhD. The collection contains 2100 images, 36 textures; 53 images per texture. Images were captured at different surface rotations:  $0^\circ, 30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ, 180^\circ$  and different camera tilt angles. The size of images is  $128 \times 128$  pixels.
- Brodatz A[29]  
This collection contains a subset of textures from Brodatz album of materials[4]. It is composed of 2048 images of 32 textures, each texture is represented with 64 images, of which 16 are original images, 16 are randomly rotated images, 16 are captured at different resolutions and 16 are randomly rotated images at various resolutions. The size of images is  $64 \times 64$  pixels.
- Brodatz B[11]  
This collection contains a subset of textures from Brodatz album of materials[4]. It is composed of 1248

images of 13 textures, each texture is represented with 96 images. Textures are captured at 6 different surface rotations:  $0^\circ, 30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ$ . The size of images is  $128 \times 128$  pixels.

- Brodatz C[5]  
This collection is also a subset of textures from Brodatz album of materials[4]. It is composed of 6720 images of 15 textures, each texture is represented with 448 images. Textures are captured at 6 different surface rotations:  $0^\circ, 30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ, 200^\circ$ . The size of images is  $32 \times 32$  pixels.

## 4.2 Analysis of the extended feature set

In this section we present results of experiments on benchmark domains, described above. The quality of image descriptions was tested by measuring the classification accuracy, obtained by algorithm SMO[21], a support vector machine variant, implemented in Weka data mining tool[32]. SMO was selected among a set of classifiers because it consistently achieved the best classification accuracy in most of the experiments. The presented results are obtained by 10-fold cross validation.

### 4.2.1 The effectiveness of additional interestingness measures

The classification accuracies, obtained by extending the default parameter set each time with one of the measures from Table 1, are presented in Table 2.

Due to unreliability of t-test[30, 6] we performed further analysis using the Friedman test[9, 10], followed with Bonferroni-Dunn test[7]. The tests confirmed that seven measures significantly improved the results, while the odds-ratio significantly decreased the accuracy. The best result was obtained with J-measure. With further experiments we tried to verify if we can obtain a better extension of the default parameter set with several parameters at once. The results were verified with Wilcoxon test[31] which revealed that adding any other measure besides J-measure does not improve the descriptions. Therefore, it is sensible to extend the default parameter set with J-measure alone.

Due to space limitation we do not show the analysis of time complexity and the numbers of generated features. The conclusions of this analysis is, that the generation of the default parameter set and the extended parameter set have approximately the same time complexity, however the number of generated features is for the extended parameter set much greater (on average for 83%), as could be expected.

### 4.2.2 The effectiveness of the multi-resolution approach

In the experiments we extended the default parameter set, with added J-measure, on original resolution  $R$  by adding

Table 2: Classification accuracies (%) of SMO on the default description versus the extended ones by various interestingness measures. The last four lines give averages and st. deviation over all datasets, the value of  $p$  for t-test and the number of wins by the extended description.

	default	$\phi$ -k.	Jacc.	conv.	gini	inter	J-meas	odds.	P.S.
outex 0	96.32	94.61	96.08	92.90	95.10	95.34	95.83	93.14	95.34
outex 1	83.04	82.45	82.99	82.06	82.84	83.63	83.53	81.86	83.38
outex 2	39.10	39.59	39.21	37.71	39.68	39.45	39.57	37.88	39.27
outex 10	93.48	94.66	93.81	94.37	94.37	94.14	94.63	94.07	93.74
outex 11	95.27	95.72	95.38	94.03	96.29	96.17	96.17	93.47	95.50
outex 12	96.62	97.50	96.64	96.81	97.19	97.12	97.19	96.64	96.66
JerryWu	86.29	98.59	87.85	98.34	97.02	84.97	96.12	95.61	93.80
brodatz B	89.75	91.57	90.16	92.23	91.15	91.40	91.40	91.40	89.50
brodatz A	78.13	84.63	79.41	85.96	82.22	85.04	82.32	82.12	79.61
brodatz C	57.36	58.20	57.71	58.05	58.16	59.12	58.41	57.23	57.72
average	81.54	83.75	81.92	83.25	83.40	82.64	83.52	82.34	82.45
st. dev.	19.05	19.60	19.03	19.84	19.46	18.84	19.45	19.57	19.23
p (t-test)		0.06	0.03	0.14	0.06	0.08	0.03	0.25	0.13
wins		8	8	6	8	8	9	5	8

parameters for resolutions  $\frac{1}{2}R$  and for resolutions  $\frac{1}{2}R + \frac{1}{4}R$  (therefore, in the latter case the feature extraction was run three times, each time on a different resolution). Table 3 gives classification accuracies of SMO classifier on all benchmark datasets for three sets of resolutions. The Friedman test shows ( $p < 0.05$ ) that the differences are significant, and the Bonferroni-Dunn test confirms that both multiresolutional descriptions (with two and three resolutions) achieved significantly better results than the single resolution description, however, the former two do not differ significantly between each other. The number of extracted features, as expected, increases approximately linearly with the number of additional resolutions, and the time complexity increases somewhat less than linearly.

#### 4.2.3 The effectiveness of meta parameters

We compared the quality of description of the default parameter set, extended with J-measure, on two resolutions ( $R$  and  $\frac{1}{2}R$ ), with the same description, extended with meta parameters, as described in Section 3. The classification accuracies for both descriptions are given in Table 4. In seven out of ten domains the meta parameters helped to improve the description, however, the Wilcoxon test ( $p < 0.05$ ) shows that overall there are no significant differences between the two descriptions. The number of features increases for 12% on average, and the average time complexity is 7 times higher for the extended description.

### 4.3 Comparison of Extended ArTex with other algorithms

We compared the Extended ArTex with several other algorithms that describe images in terms of a set of numerical

Table 3: Classification accuracy (%) of SMO using the default parameter set, with J-measure included, when generated on one resolution  $R$ , two resolutions  $R + \frac{1}{2}R$ , and three resolutions  $R + \frac{1}{4}R + \frac{1}{8}R$ . The last four lines have the same interpretation as in Table 2.

	$R$	$R + \frac{1}{2}R$	$R + \frac{1}{2}R + \frac{1}{4}R$
outex 0	95.83	98.05	95.11
outex 1	83.53	83.97	81.32
outex 2	39.57	52.26	54.78
outex 10	94.63	98.38	98.26
outex 11	96.17	97.07	95.84
outex 12	97.19	97.95	97.84
JerryWu	96.12	99.40	99.75
brodatz B	91.40	97.60	97.60
brodatz A	82.32	94.01	94.77
brodatz C	58.41	60.93	58.41
average	83.52	87.96	87.37
st. dev.	19.45	17.24	17.03
p (t-test)		0.01	0.03
wins		10	7

features. We selected the following algorithms: a similar algorithm to ArTex, also based on association rules, developed by Rushing et al.[23], second order statistics as a standard benchmark algorithm[12] and the three advanced approaches which are effective and also often applied[13]: Laws filters[18], Haar waves, and Gabor waves[19]. Table 5 gives the classification accuracies of SMO algorithm using the 10-fold cross-validation.

The Friedman test ( $p < 0.05$ ) followed by the Bonferroni-Dunn test ( $p < 0.05$ ), shows that the Extended

Table 4: Classification accuracy (%) of SMO using the default parameter set, with included J-measure, on two resolutions  $R + \frac{1}{2}R$ , without and with added meta-parameters. The last four lines have the same interpretation as in Table 2. Columns  $d$  in  $|d|$  represent the difference of classification accuracies and its absolute value, and the last column contains ranks for  $|d|$ , used by the Wilcoxon test.

	without meta p.	with meta p.	d	d	rank
outex 0	98.05	97.92	-0.13	0.13	2
outex 1	83.97	87.71	+3.74	3.74	9
outex 2	52.26	55.00	+2.74	2.74	7
outex 10	98.38	98.54	+0.16	0.16	3
outex 11	97.07	97.50	+0.43	0.43	4
outex 12	97.95	98.96	+1.01	1.01	6
JerryWu	99.40	99.41	+0.01	0.01	1
brodatz B	97.60	96.84	-0.76	0.76	5
brodatz A	94.01	97.19	+3.18	3.18	8
brodatz C	60.93	53.12	-7.81	7.81	10
average	87.96	88.22			
st. dev.	17.24	18.31			
p (t-test)		0.40			
wins		7			

ArTex is significantly better than the second order statistics, while there are no significant differences between ArTex and other algorithms. This confirms that the Extended ArTex achieves the performance of best algorithms for generating texture descriptions. With respect to the time complexity, the most time consuming algorithms are Gabor waves, Rushing and Extended ArTex. When compared with Extended ArTex, the Gabor waves algorithm is on average 141 times slower, the algorithm by Rushing et al. has on average the same time complexity, while the Laws filters are 2 times faster, the second order statistics 10 times faster and the Haar waves 15 times faster.

Table 5 indicates that the Extended ArTex achieves best results in domains with relatively large images ( $128 \times 128$  or more), and performs relatively poorly on domains with small images, like *Outex 2* and *Brodatz C*. This is in accordance with our expectations. For reliable description of images, ArTex requires a large texture in order to be able to reliably evaluate the values of parameters - statistical estimates - whose accuracy depends on the number of pixels.

## 5 The performance of Extended ArTex in several real-world applications

In this section we provide the performance comparison with other feature extraction algorithms in some applications which were implemented at our institution by using the Extended ArTex. We briefly outline the real-world data sets and then we provide the results.

We applied our approach on four real-world problems:

- *pH6pH10*  
The problem is to differentiate microscopic pictures of 0.1% suspension of  $Al_2O_3$  in distilled water ( $H_2O$ ) with two different acidities: pH6 and pH10. The pictures were prepared at the University of Stuttgart[16, 17]. The size of images is  $300 \times 300$  pixels. There are 30 images for each of the two classes.
- *Materials*  
Each of six different materials was photographed 50 times on various places. The size of images is  $200 \times 200$ . Therefore this domain consists of six classes, where each class contains 50 samples. The task is to classify new images into one of six classes.
- *SyringeStactometer*  
This domain consists of two classes of microscopic images of dried drops of tap water, that have to be differentiated. The first class of 52 drops was created using a syringe, whereas the second class of 61 drops was created using a stactometer. The size of images is  $640 \times 640$ .
- *Coronas*  
This domain consists of GDV images of human fingertips. GDV images are obtained from BEO GDV Camera, developed by Korotkov[15]. The camera captures  $320 \times 240$  gray scale images. The first class of 289 images consists of fingertips of humans in normal state, and the second class of 217 images consists of fingertips of humans in the altered state of consciousness. Most of images were obtained from Technical University SPIFMO in St. Petersburg, Russia. The task is to automatically detect the altered state of human



Table 5: Classification accuracies (%) of SMO using descriptions of various feature extraction algorithms. The last four lines have the same interpretation as in Table 2.

	Ext.ArTex	Laws	Haar	2 <sup>nd</sup> Stats	Gabor	Rushing
outex 0	97.92	91.25	93.13	72.29	99.38	97.07
outex 1	87.71	90.25	89.39	71.73	98.30	87.40
outex 2	55.00	84.59	72.09	59.70	94.49	24.35
outex 10	98.54	83.96	86.94	82.62	97.80	86.16
outex 11	97.50	92.19	93.75	75.63	99.79	97.97
outex 12	98.96	82.72	87.81	76.98	96.83	86.23
JerryWu	99.41	22.95	83.19	75.86	53.76	83.51
brodatz B	96.84	78.21	94.07	96.00	86.70	69.15
brodatz A	97.19	62.30	65.43	18.31	87.89	12.96
brodatz C	53.12	66.96	65.43	48.88	70.77	11.73
average	88.22	75.54	83.12	67.80	88.57	65.65
st. dev.	18.31	20.99	11.34	21.43	15.12	35.07
p (t-test)		0.10	0.14	0.01	0.47	0.01
wins		3	3	1	5	2

consciousness, using the GDV camera.

Table 6 gives the classification accuracies of SMO obtained by 10-fold cross-validation on the described datasets. The best results were achieved by the Extended ArTex, except in domain Coronas, where the second order statistics perform the best.

## 6 Conclusions

We described the extended textural features which are based on association rules. We extended the basic textural features of our algorithm ArTex[2] in three ways: by using various interestingness measures, by using the multi-resolution approach, and by using the meta-parameters.

Among various interestingness measures, the most promising results were achieved with J-measure. J-measure is, as far as we know, the best measure for evaluating the quality of decision (if-then) rules, developed by Smyth and Goodman back in 1990[24]. It has nice properties, such as nonnegativity, transparent interpretation (the sum of J-measures over all possible values is equal to the information gain[14]), it is possible to estimate the upper bound, and it is approximately  $\chi^2$  distributed. Therefore, it is not surprising, that it achieves the best results in our experiments.

Results from our experiments show that the multi-resolution approach enables to obtain significantly better descriptions of images. For texture descriptions with association rules this could be expected. Not every combination of scale and neighborhood size can guarantee that the pattern would be detected. As illustrated by Figures 1 and 2, the effectiveness of descriptions, which use an upper bound for the neighborhood around the central pixel, highly depends on the scale. As it is not known in advance

which resolution will provide the best results, our approach combines the features, obtained from several resolutions. In our current research we try to develop an efficient algorithm for automatic detection of a small subset of relevant resolutions.

Our definition of meta-parameters is the first step towards the ultimate goal, which we hope to reach in further work: to get higher order association rules, which would capture the global structure of the image and would also allow for transparent (human readable) description of the images. The results, presented in this paper, are promising and have encouraged us to continue with research in this direction.

Overall, the extended representation improves the utility of basic textural features and often gives better results with respect to basic features, derived from association rules, as well as with respect to standard texture descriptions. The results in real-world applications prove the utility of our approach.

Our current research is devoted to describe medical images in order to obtain reliable diagnostic rules. The preliminary experiments in two medical domains are quite promising. In diagnosing the coronary artery disease from myocardial perfusion scintigraphy images the preliminary results show the significant improvement over the accuracy of physicians experts. In another diagnostic problem, based on the whole-body bone scintigrams, the picture is not so clear, still however, the Extended ArTex achieves significantly better results than any other approach to feature extraction.

Yet another line of research is the visualization of textures from their association rule descriptions. As the number of extracted features may be relatively large (typically several hundreds of features are generated in each problem domain), such amount of features cannot be transparent to

Table 6: Classification accuracies (%) on real-world data sets of SMO using descriptions of various feature extraction algorithms. The last four lines have the same interpretation as in Table 2.

	Ext.ArTex	Laws	Haar	2 <sup>nd</sup> Stat.	Gabor	Rushing
pH6pH10	100.00	100.00	100.00	100.00	100.00	100.00
Materials	100.00	97.00	99.67	99.00	100.00	100.00
Syrin.Stact.	89.91	87.42	86.74	85.96	85.83	80.73
Coronas	82.80	80.25	80.23	84.78	78.81	70.00
average	93.18	91.17	91.66	92.44	91.16	87.68
st. dev.	8.40	9.04	9.81	8.18	10.60	14.88
p (t-test)		0.03	0.08	0.30	0.09	0.10
wins		0	0	1	0	0

a human expert. To overcome this we plan to develop a visualization tool for a user-friendly presentation of derived (large) sets of features in the form of artificial textures. The preliminary results in this area indicate that users may find on such artificially generated textures important patterns, related to the target concepts, and that users may be able to notice also more or less obvious differences in artificially generated textures for different classes of images.

## References

- [1] R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. In P. Buneman and S. Jajodia, editors, *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, pages 207–216, Washington, D.C., 26–28 1993.
- [2] M. Bevk and I. Kononenko. Towards symbolic mining of images with association rules: Preliminary results on textures. *Intelligent Data Analysis*, 10(4):379–393, 2006.
- [3] S. Brin, R. Motwani, and C. Silverstein. Beyond market baskets: Generalizing association rules to correlations. In J. Peckham, editor, *SIGMOD 1997, Proceedings ACM SIGMOD International Conference on Management of Data, May 13-15, 1997, Tucson, Arizona, USA*, pages 265–276. ACM Press, 1997.
- [4] P. Brodatz. *Textures - A Photographic Album for Artists and Designers*. Reinhold, Dover, New York, 1966.
- [5] J. M. Carstensen. Cooccurrence feature performance in texture classification. In *The 8th Scandinavian Conference on Image Analysis, Tromsø, Norway*, pages 831–838, may 1993.
- [6] J. Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30, 2006.
- [7] O. J. Dunn. Multiple comparisons among means. *Journal of the American Statistical Association*, 56:52–64, 1961.
- [8] P. A. Flach and N. Lachiche. Confirmation-guided discovery of first-order rules with tertius. *Machine Learning*, 42(1/2):61–95, 2001.
- [9] M. Friedman. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 2:675–701, 1937.
- [10] M. Friedman. A comparison of alternative tests of significance for the problem of m rankings. *Annals of Mathematical Statistics*, 11:86–92, 1940.
- [11] G. M. Haley and B. S. Manjunath. Rotation-invariant texture classification using a complete space-frequency model. *IEEE Tr. Im. Proc.*, 8(2):255–269, Feb. 1999.
- [12] R. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 3(11):610–621, 1973.
- [13] P. E. T. Jorgensen. *Analysis and probability: wavelets, signals, fractals*, volume 234 of Graduate Texts in Mathematics. Springer, New York, 2006.
- [14] I. Kononenko and M. Kukar. *Machine Learning and Data Mining: Introduction to Principles and Algorithms*. Horwood Publ., 2007.
- [15] K. Korotkov. *Aura and Consciousness*. State Editing & Publishing Unit “Kultura”, 1998. St.Petersburg, Russia.
- [16] B. Kröplin, M. Hein, J. Herrmann, and B. Heusel. Report on the microoptic investigations of water and watery solutions. Technical report, University of Stuttgart, Institute for Statics and Dynamics of Aerospace Structures, 05-2000.

- [17] B. Kröplin, M. Hein, J. Herrmann, and B. Heusel. Mikrooptische Untersuchungen zum Einfluss von elektromagnetischen Feldern, Magneten und Handys auf Wasser. Technical report, Universität Stuttgart, Institut für Statik und Dynamik der Luft und Raumfahrtkonstruktionen, 06-2000.
- [18] K. I. Laws. *Textured image segmentation*. PhD thesis, Dept. Electrical Engineering, University of Southern California, 1980.
- [19] B. S. Manjunath and W. Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(8):837–842, 1996.
- [20] T. Ojala, T. Mäenpää, M. Pietikäinen, J. Viertola, J. Kyllönen, and S. Huovinen. Outex- new framework for empirical evaluation of texture analysis algorithms. In *ICPR (1)*, pages 701–706, 2002.
- [21] J. C. Platt. Fast training of support vector machines using sequential minimal optimization. In *Advances in kernel methods: support vector learning*, pages 185–208. MIT Press, 1999.
- [22] A. Rosenfeld and A. Kak. *Digital Picture Processing (2nd Edition)*, volume 2. Academic Press, Inc., 1982.
- [23] J. A. Rushing, H. S. Ranganath, T. H. Hinke, and S. J. Graves. Using association rules as texture features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(8):845–858, 2001.
- [24] P. Smyth and R. M. Goodman. Rule induction using information theory. In G. Piatetsky-Shapiro and W. Frawley, editors, *Knowledge Discovery in Databases*. MIT Press, 1990.
- [25] P. Tan and V. Kumar. Interestingness measures for association patterns: A perspective. Technical Report TR00-036, Department of Computer Science, University of Minnesota, 2000.
- [26] P. Tan, V. Kumar, and J. Srivastava. Selecting the right interestingness measure for association patterns. In *Proceedings of the Eight ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2002.
- [27] M. Tuceryan and A. Jain. Texture segmentation using voronoi polygons. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 211–216, 1990.
- [28] M. Tuceryan and A. Jain. *The Handbook of Pattern Recognition and Computer Vision (2nd Edition)*. World Scientific Publishing Co., 1998.
- [29] K. Valkealahti and E. Oja. Reduced multidimensional cooccurrence histograms in texture classification. *PAMI*, 20(1):90–94, January 1998.
- [30] G. I. Webb. MultiBoosting: A technique for combining Boosting and Wagging. *Machine Learning*, 40(2):159–196, 2000.
- [31] F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics*, 1, 1945.
- [32] I. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann, 1999.
- [33] J. Wu. *Rotation Invariant Classification of 3D Surface Texture Using Photometric Stereo*. PhD thesis, Heriot-Watt University, Edinburgh, 2003.

