

# Emergent Semantic Patterns In Large Scale Image Dataset: A Datamining Approach

Umair Mateen Khan  
Computer Science Department  
Otago University  
Dunedin, New Zealand  
Email: umairkhan@cs.otago.ac.nz

Brendan McCane  
Computer Science Department  
Otago University  
Dunedin, New Zealand  
Email: mccane@cs.otago.ac.nz

Andrew Trotman  
Computer Science Department  
Otago University  
Dunedin, New Zealand  
Email: andrew@cs.otago.ac.nz

**Abstract**— In this paper we investigate an unsupervised learning method applied to low level image features extracted from a large collection of images using data mining strategies. The mining process resulted in several interesting emergent semantic patterns. Initially, local image features are extracted using image processing techniques which are then clustered to generate a bag of words (BoW) for each image. These bags of words are then used for mining co-occurring patterns. The generated patterns were either global in nature i.e showed a behavior spread across many images or a local and more rare behavior found across few images. These patterns are assigned semantic names to build a semantic relationship among images containing them.

## I. INTRODUCTION

There has recently been a strong interest in the learning of low-level features from training data, with a focus on neurally inspired architectures [1], [2], [3], [4]. At the same time, the number of large image databases has increased [5], [6]. In this paper we hypothesize that investigating very large image databases will result in the emergence of interesting patterns in a bottom-up sense. However, we do not start at the lowest possible level (pixels), and instead start at an intermediate level (SIFT features [7]), and investigate the co-occurrence of such features to see if more semantic level features emerge. We present an approach that explores emergent patterns in a large image collection by using both frequent and rare itemset mining strategies.

### *Related Work*

Several significant works have looked at the application of data-mining techniques for image retrieval and object recognition applications. The most common method used is association rule mining. Association rule mining is a technique for discovering both global and local knowledge from large collections of data [8]. Data mining strategies like Apriori [9], Eclat [10] and Fp-Growth [11] have been used for finding correlations among data elements and extracting hidden relationships that are not so obvious. These techniques can be applied to any field that produces or deals with a large amount of data. Some of the applications of data mining include market basket data analysis, patient disease symptoms analysis, stock analysis and social network analysis. Similarly image mining is no exception and deals with discovering hidden relationships from image data elements that are pixels,

shapes, textures or higher level features such as SIFT [7]. For all the above mentioned mining techniques and many others it is necessary for the data to be represented as market basket transactions [12] in which each transaction can contain 1 or multiple items.

Association mining among elements of data is a highly researched area but rather less studied in the context of image mining [13]. In [14] an attempt was made to mine relationships among objects from different modalities of multimedia data. These objects were called perceptual objects and were defined in a spatio-temporal window and the generated rules were used to define the relationships in a more compact and semantic way. Another approach to mine frequently co-occurring objects/actors or scenes in video frames was performed in [15]. In their approach they created a transaction for each visual word and all its neighboring words were taken as its items. Once these neighborhoods are defined they are used for mining co-occurring objects or actors among many frames. Association rule mining has also been used for clustering web images [16]. The authors first generated rules using both visual and text features obtained from web pages and then used them to generate hyper-graphs. Then a hyper-graph partitioning algorithm was used to get clusters. In [13] association rule mining was performed on regions of interest (ROI) in CT images of brains. The ROI were first extracted using a region extraction and clustering algorithm which also used domain knowledge. The rules were then generated on mined frequent itemsets considering ROI of brain images as items. In another approach [17], knowledge discovery from image contents without using any domain knowledge was investigated. Their work was based on finding relationships in basic geometric shapes but was performed on a very limited number of images. They also suggested the need for some reliable domain knowledge for better object recognition. Object class recognition was also done using association rule mining techniques in [18]. The association rules were built on low level features occurring within a bounding box containing either a background image example or an object image from one of its classes. Visual words inside this box were represented as transactions and then association rules were mined. The transaction database contained a combined set of transactions obtained from both objects and background bounding boxes. The learned rules

were then used to tell the presence of a particular object class or a background in the unseen images. A similar approach was presented in [19] where they used it for detecting logos of different categories in an image. They achieved this by locating dense configurations of frequent local features related to each logo class. They extracted association rules on a spatial pyramid of each base feature where a base feature is represented as a group of all neighboring features inside a radius grid across it. Each base feature was then represented as a transaction and all surrounding features as items of the transaction. Human action classification by mining association rules was performed in [20]. The key idea was in the concept of compound features. Compound features were defined as groups of corner descriptors used to encode local features in space and time. These features were learned using data mining techniques by looking at their co-occurrences. The classifier was actually a group of these computed features and was capable of both recognizing and localizing a real time activity.

In all previous work the focus of using data mining techniques on images has been on a specific application. Those applications include: object recognition; object classification; clustering; scene recognition; or content-based image retrieval. The purpose of the work presented here is quite different, although the techniques used are similar. Our goal is to show that semantic level features of images can emerge in a bottom-up process from a suitably large collection of images, and that these semantic features are interesting in themselves. Our long term goal is to show that the notion of semantic-ness is crucially dependent on the low-level processing of a particular vision system (natural or artificial) and that different vision systems will give rise to different notions of “semantic” and will necessarily interpret the world in different, but equally valid, ways.

## II. FEATURE EXTRACTION AND CLUSTERING

The current generation of object recognition, object classification and image retrieval systems identify certain interest points, called features using an image processing technique such as SIFT [7] or SURF [21]. To extract these features, techniques such as edge detection, corner detection [22], blob [23] and ridge detection [24] etc. are used. In this paper SIFT [7] is used to identify keypoints and the 128 dimensional SIFT descriptor is used to represent the keypoint. This descriptor contains orientation information obtained from a 4x4 grid across each keypoint. SIFT features are highly distinctive and provide invariance to different transformations e.g. rotation, translation and scaling and partial invariance to illumination and viewpoint change. SIFT has been extensively used in many object recognition and image retrieval applications [7], [18], [25], [26], [27], [28]. SIFT produces a large number of features and may generate 1000 to 2000 features per image depending on image content and size. SIFT features are clustered into a finite set of visual words using approximate K-means clustering as done in [26], [27]. For fast lookup of nearest clusters a Kd-Tree [29] is built on cluster centers. For clustering we partitioned the data into chunks so that it can be loaded in

memory. Each image is represented using a visual bag of words [26] representation and the co-occurrence relationships between the words are mined to extract interesting features as outlined in the following section.

## III. FREQUENT ITEMSETS AND ASSOCIATION RULES MINING

Association rule mining is a technique used for mining interesting relationships in market basket transaction data. A market basket is a list of items that were purchased together by a customer. The association found is not influenced by the inherent property of the data but rather the co-occurrences in data items. Association rule mining was first suggested in [12].

Let  $I = \{i_1, i_2, i_3, \dots, i_n\}$  be the set of all possible items in the data collection (all possible visual words in our case). Let  $T$  be a transaction containing a subset of these items such that  $T \subseteq I$ . In our case,  $T$  contains all visual words from a single image.  $D$  is the collection of all transactions. An association rule is an implication,  $X \Rightarrow Y$ , where  $X \subset I$  and  $Y \subset I$  but  $X \cap Y = \emptyset$  [12]. Association rules have two parameters called *support*  $s$  and *confidence*  $c$ . Support is a count of transactions in which a particular item is found. So the support of an association rule  $X \Rightarrow Y$  which contains two items would be the ratio of transactions that contains  $X \cup Y$  compared to the total number of transactions. Confidence is the ratio of the number of transactions that contain  $X \cup Y$  to the total number of transactions that contain  $X$ . An association rule only holds if it has a support  $s$  greater than *minsup* and confidence  $c$  greater than *minconf*; values which are specified by the user.

Apart from the *minconf* and *minsup* values a further measure called *lift* can also be used to check the strength of a rule. The lift value of a rule tells the degree to which  $Y$  is more likely to occur when  $X$  occurs. A value of lift less than 1.0 means that  $Y$  is less likely to occur with  $X$  than  $Y$ 's total support in the entire transaction dataset. A lift value higher than 1.0 implies a positive association between  $X$  and  $Y$ . A positive association means that  $X$  and  $Y$  co-occur together more than expected.

The *support* of a rule can be defined as:

$$\text{support}(X \Rightarrow Y) = \frac{\#\_transactions\_containing\_both\_X\_and\_Y}{total\_#\_of\_transactions} \quad (1)$$

The *confidence* for the rule can be calculated as:

$$\text{confidence}(X \Rightarrow Y) = \frac{\#\_transactions\_containing\_both\_X\_and\_Y}{\#\_of\_transactions\_containing\_X} \quad (2)$$

And the *Lift* value of a rule can be defined as:

$$\text{lift}(X \Rightarrow Y) = \frac{\text{confidence}(X \Rightarrow Y)}{\text{support}(Y)} \quad (3)$$

The process of association rule generation is a two stage process. First frequently co-occurring items called *itemsets*

are mined using a *minsup* threshold. An itemset containing 'r' items is called an *r-itemset*. Now association rules are generated from these itemsets and *minconf* and *lift* thresholds are checked. The current BOW representation of all images can be converted to market basket transactions by seeing each image as a transaction and words in that image as items of that transaction. Co-occurrence among these transactions can be found using any such association rule mining algorithm. We used the *Eclat* [10] algorithm for mining these co-occurring itemsets or frequent itemsets from image transactions. Although *Fp-Growth* [11] yielded much better results to *Eclat* as described in [30], we used *Eclat* because in our case *Fp-Tree* did not fit in memory. Once we have the generated itemsets we then generated the association rules and only those rules are considered which met the *minconf* and *lift* criteria.

#### IV. RARE ITEMSETS AND ASSOCIATION RULES MINING

Mining frequent patterns from the data is crucial and it gives a global insight into the data. But in some cases that global insight can be easily predicted by domain experts and hence does not necessarily give useful knowledge. For example if we look at the records of patient history for a fatal disease then common symptoms can be easily mined and most of them would likely be already known to domain experts. In that case a more interesting finding would be to see which symptoms occurred rarely or were infrequent but had very high confidence. A frequent itemset miner completely ignores such itemsets because they occur in very few transactions.

An obvious way to find such rare co-occurrences is to reduce the *minsup* to a very low value and then a frequent itemset miner will consider these rare occurrences as frequent. This results in a very high running time of the algorithm and too many itemsets fulfilling the *minsup* threshold, this phenomena is known as the *rare itemset problem* [31]. The algorithms that are designed to mine rare itemsets use different notions of such thresholds.

For our problem we used the *RP-Tree* [31] algorithm that extracts rare patterns by building a prefix tree only on those transactions that contain a rare item. The algorithm is a modification of the *Fp-Growth* [11] algorithm. Two thresholds are used for mining rare itemsets, the first called *minRareSup*, is a minimum support for an item to be a rare item and works as a noise filter. All those items having support less than this threshold are not considered further. The second threshold *minFreqSup* is a maximum support for an itemset to be considered as a rare itemset [31]. All items having support greater than this threshold are considered frequent.

Itemsets are categorized into three different types [31]: the first is *rare itemset* and it includes all those itemsets that have support less than *minFreqSup* but greater than or equal to *minRareSup*. The *rare-item itemsets* consists of those itemsets that have both rare and frequent items but the itemset itself is rare and therefore it also fulfills the criteria for a *rare itemset*. The third type is *non-rare-item itemset* and consists of all the items which are frequent but the itemset itself is rare. We are

mainly concerned with first two types of itemsets because non-rare-item itemsets are very likely to occur by chance in large datasets.

Consider the itemset  $X$ . It is called a *rare itemset* iff

$$\begin{aligned} \text{support}(X) < \text{minFreqSup}, \\ \text{support}(X) \geq \text{minRareSup} \end{aligned} \quad (4)$$

$X$  is called *rare-item itemset* iff

$$\begin{aligned} \exists x \in X, \text{support}(x) < \text{minFreqSup}, \\ \text{support}(X) < \text{minFreqSup} \end{aligned} \quad (5)$$

$X$  is called a *non-rare-item itemset* iff

$$\begin{aligned} \forall x \in X, \text{support}(x) \geq \text{minFreqSup}, \\ \text{support}(X) < \text{minFreqSup} \end{aligned} \quad (6)$$

#### V. EXPERIMENTAL SETUP AND RESULTS

The focus of this research is to investigate emergent patterns from a large image collection. The *MIRFLICKR* [32] collection containing 1 million images was downloaded. We used half of the images because we considered it to be sufficient for our experiments. The dataset contains high quality images of everyday scenes and is designed for image retrieval applications. It has been used in *ImageCLEF* [6] evaluation forum for the last 3 years. Some of the images from this collection are shown in Figure 1. The dataset was suited to our scenario because we are learning unsupervised, and do not want to use a dataset containing many similar scenes. For feature extraction, *SIFT* [7] was used and more than 10 billion local features were extracted. Features from each image were saved in separate files along with their X and Y coordinates. All experiments were performed on 64-bit Intel Core 2 Duo 3.00 GHz CPU with internal memory of 4 GB.

Approximate k-means clustering was then performed on all the features and an image chunk size of 10,000 images was chosen because that was the largest chunk size that could consistently fit all of the features in memory. For one iteration of the clustering process features from 50 chunks were processed i.e a total 500,000 images. The clustering process was terminated after 10 such iterations and the clustering took about 12 hours.

The selection of  $k$  (total cluster centers) was very important and the most crucial in our case as a wrong value of  $k$  can greatly effect the mining process by increasing or decreasing the co-occurrences. A large value of  $k$  can increase false negatives as most features will match to different clusters even though there would be only slight differences. On the other hand a small value of  $k$  will increase false positives as many features that are different from each other will match the same cluster center. As we were unsure about the correct number of clusters we decided to use five different values of  $k$ . These values were set to 5,000, 15,000, 35,000, 50,000 and 75,000 respectively. Once these clusters were obtained from the clustering process, all images were then represented in their BoW form.

After applying frequent itemset mining we observed images contained in these frequent itemsets. We found the value of  $k$  to work well between 35,000 and 50,000 cluster centers because most of the patterns that emerged, were observed with these two clusters. The  $minsup$  threshold was set at 0.025% and 0.05% for mining frequent itemsets for all cases because a lower threshold generated too many itemsets while a higher value resulted in very few itemsets as explained in the next section. For rare itemset mining we only used the transaction datasets generated from 35,000 and 50,000 cluster centers. The  $minFreqSup$  was chosen to be 0.04%, 0.05% and 0.06% and  $minRareSup$  was selected as 0.002%, 0.004% and 0.006%, which appeared to be high enough to differentiate noise from a real rare occurrence. The association rules were also generated from these itemsets and only rules with  $confidence \geq 0.9$  and  $lift \geq 1.0$  were considered interesting.

Here, it is worth noting that although we generate association rules for both mining processes we did not use them for defining relationships among items or image words. Rather we used them for selecting interesting itemsets to test the hypothesis that interesting semantic features will emerge from a large collection of data.

#### A. Frequent Patterns and Generated Rules

The mining process generated a large number of itemsets e.g. in case of 35,000 clusters up to 1 million itemsets were generated as shown in Table I. Because of this large number it was not possible for us to view images generated by all these itemsets and then decide whether it results in interesting information or not. Instead, we generated association rules with a very high confidence value and used it to prune uninteresting itemsets. Only those itemsets were chosen for viewing their associated images for which there existed a rule that satisfied this confidence criteria. Doing this greatly reduced the total number of itemsets to be viewed. But still in some cases the remaining number of itemsets was very high leaving us with no other option than to randomly sample itemsets for viewing. The sampling was done by keeping the same proportion of itemsets to be examined from varying itemset lengths. On observing images containing these itemsets we found some interesting patterns showing global behavior of these images. From the total itemsets resulted from each mining process approximately 200 itemsets were viewed and each of these were manually categorised into six semantic categories.

The semantic categories were: stripes or parallel lines, dots and checks, bright dots, single lines, intersections and frames as shown in Figure 2. For example all images containing a "dots and checks" pattern had the same semantic concept identified by the itemset (a set of SIFT features in the image) as shown in Figure 2(b). The red marks are the features associated with items (words) in this itemset. Images in each category of Figure 2 were chosen randomly from all the images containing that itemset.

Apart from these patterns, we observed some other patterns showing different behaviors e.g. there were some itemsets

Table I  
NUMBER OF FREQUENT ITEMSETS GENERATED AGAINST  $k$  CLUSTER CENTERS USING TWO DIFFERENT MIN.SUPPORT THRESHOLDS. THE MINING PROCESS WAS ABORTED WHEN THE GENERATED FILE SIZE REACHED TO 10GB.

$MinSup$	Total number of itemsets for different cluster centers				
	5,000	15,000	35,000	50,000	75,000
0.025%	aborted	aborted	988,354	427,398	86,203
0.05%	aborted	1,085,926	32,852	14,754	3,129

Table II  
ASSOCIATION RULES GENERATED FROM FREQUENT ITEMSETS FOR TWO VALUE OF CLUSTER CENTERS AND TWO MIN.SUPPORT THRESHOLDS.

$MinSup$	Total number of rules for different cluster centers	
	35,000	50,000
0.025%	71,645	2,598
0.05%	752	556

behaving like an efficient text detector while others resulted in images having either black frames, double frames or hollow circles. But for these itemsets there was also an overlap of images from one or more different semantic patterns e.g. the text generating pattern contained images of stripes as well. So we did not categorize them as a separate class rather kept them in one of these 6 categories. Figures 2(a) and 2(f) show some images containing text and different types of framed images. These images were classified into the main categories of stripes and frames respectively. It is also interesting to note that with a different number of clusters not all of these patterns were identified which clearly demonstrates the effect of number of clusters on co-occurrence among features. For the case of 50,000 clusters with  $minSup$  threshold of 0.025% we only observed patterns a,b,c and f as depicted in Figure 2.

From Table I we can also see the effect of number of clusters on total number of itemsets. Fewer cluster centers generated too many co-occurrences and hence too many itemsets. In some cases we aborted the mining either because of too much time spent by the mining process or the generated file size was too large. A higher value of  $k$  on the other hand generated fewer co-occurrences and hence considerably fewer itemsets. As neither too many or too few co-occurrences were desirable, the values of  $k$  of 35,000 and 50,000 appeared reasonable.

We observed that with  $minSup$  of 0.05% the number of rules generated was much less with both number of clusters. When viewing images generated by these itemsets we only observed three categories. Patterns a, b and f from Figure 2 were detected for 35,000 clusters while for 50,000 clusters the detected patterns were a, b and c. We found that because of a higher  $confidence$  (i.e 90%) most of the itemsets belonging to other categories were pruned. The other patterns were detected when this threshold was reduced to 60%.

For the selection of interesting itemsets we extracted association rules only from itemsets generated from 35,000 and 50,000 cluster centers. Table II depicts the total rules generated against each number of cluster centers.



Table IV

NUMBER OF ASSOCIATION RULES GENERATED FOR RARE ITEMSETS FOR 35,000 AND 50,000 CLUSTER CENTERS. TWO *minRareSup* THRESHOLDS AND THREE DIFFERENT *minFreqSup* WERE USED. THE EMPTY CELL SHOWS THAT NO RULES WERE FOUND BECAUSE OF VERY LESS NUMBER OF ITEMSETS.

<i>MinFreqSup</i>	35,000		50,000	
	0.004%	0.006%	0.004%	0.006%
0.04%	-	-	49,515	52,243
0.05%	1,247	215	11,758	8,253
0.06%	9,616	14,120	26,121	78,417

frames. We had hoped that some notion of “object-ness” might have emerged from the data, but without using spatial information of the features, this hope was forlorn. Instead, non-local semantic features emerged. An obvious extension is to investigate the effect of also using spatial information when generating interesting rules. Further work on determining whether these bottom-up semantic elements are useful for any of the typical image classification tasks is also warranted.

#### ACKNOWLEDGMENT

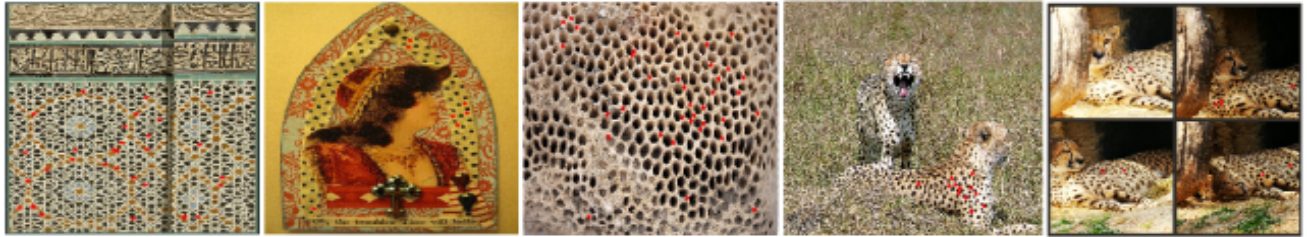
The authors would like to thank Christian Borgelt and Yun Sing Koh for providing their implementation code for *Eclat* and *Rp-Tree* algorithms respectively.

#### REFERENCES

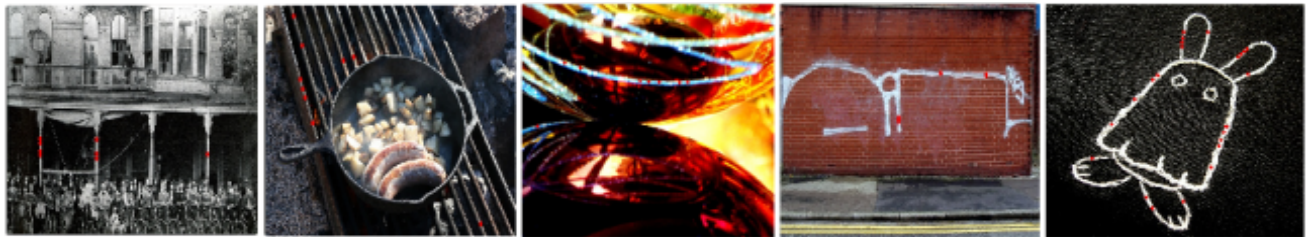
- [1] Q. Le, J. Ngiam, Z. Chen, D. Chia, P. Koh, and A. Ng, “Tiled convolutional neural networks,” *Advances in Neural Information Processing Systems*, vol. 23, 2010.
- [2] H. Lee, R. Grosse, R. Ranganath, and A. Ng, “Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations,” in *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, 2009, pp. 609–616.
- [3] Y. LeCun, K. Kavukcuoglu, and C. Farabet, “Convolutional networks and applications in vision,” in *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*. IEEE, 2010, pp. 253–256.
- [4] A. Coates, H. Lee, and A. Ng, “An analysis of single-layer networks in unsupervised feature learning,” *JMLR Workshop and Conference Proceedings*.
- [5] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A Large-Scale Hierarchical Image Database,” in *CVPR09*, 2009.
- [6] M. Sanderson and P. Clough, “ImageCLEF-Cross Language Image Retrieval Track,” <http://http://imageclef.org/>.
- [7] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” vol. 60, no. 2, 2004, pp. 91–110.
- [8] U. Fayyad, G. Piatetsky-shapiro, and P. Smyth, “From data mining to knowledge discovery in databases,” vol. 17, 1996, pp. 37–54.
- [9] R. Agrawal and R. Srikant, “Fast algorithms for mining association rules in large databases,” in *VLDB’94, Proceedings of 20th International Conference on Very Large Data Bases, September 12-15, 1994, Santiago de Chile, Chile*, J. B. Bocca, M. Jarke, and C. Zaniolo, Eds. Morgan Kaufmann, 1994, pp. 487–499.
- [10] M. J. Zaki, S. Parthasarathy, M. Ogihara, and W. Li, “New algorithms for fast discovery of association rules,” in *3rd International Conference on Knowledge Discovery and Data Mining (KDD)*, Aug 1997.
- [11] J. Han, J. Pei, Y. Yin, and R. Mao, “Mining frequent patterns without candidate generation: A frequent-pattern tree approach,” vol. 8, no. 1, 2004, pp. 53–87.
- [12] R. Agrawal, T. Imielinski, and A. N. Swami, “Mining association rules between sets of items in large databases,” in *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, D.C., May 26-28, 1993*, P. Buneman and S. Jajodia, Eds. ACM Press, 1993, pp. 207–216.
- [13] H. Pan, Q. Han, G. Yin, W. Zhang, J. Li, and J. Ni, “A roi-based mining method with medical domain knowledge guidance,” in *Proceedings of the 2008 International Conference on Internet Computing in Science and Engineering*, ser. ICICSE ’08. Washington, DC, USA: IEEE Computer Society, 2008, pp. 91–97. [Online]. Available: <http://dx.doi.org/10.1109/ICICSE.2008.91>
- [14] J. Martinet and S. Satoh, “A study of intra-modal association rules for visual modality representation,” in *Content-Based Multimedia Indexing, 2007. CBMI’07. International Workshop on*. IEEE, 2007, pp. 344–350.
- [15] T. Quack, V. Ferrari, and L. V. Gool, “Gool. video mining with frequent itemset configurations,” in *In Proc. CIVR*. Springer, 2006, pp. 360–369.
- [16] H. H. Malik, “Clustering web images using association rules, interestingness measures, and hypergraph partitions,” in *ICWE 06: Proceedings of the 6th international conference on Web engineering*. ACM Press, 2006, pp. 48–55.
- [17] C. Ordóñez and E. Omiecinski, “Discovering association rules based on image content,” in *Proceedings of the IEEE Advances in Digital Libraries Conference (ADL’99, 1999*, pp. 38–49.
- [18] T. Quack, V. Ferrari, B. Leibe, and L. J. V. Gool, “Efficient mining of frequent and distinctive feature configurations,” in *ICCV*, 2007, pp. 1–8.
- [19] J. Kleban, X. Xie, and W.-Y. Ma, “Spatial pyramid mining for logo detection in natural scenes,” in *ICME*, 2008, pp. 1077–1080.
- [20] A. Gilbert, J. Illingworth, R. Bowden, and G. X. England, “Scale invariant action recognition using compound features mined from dense spatiotemporal corners,” in *In ECCV*, 2008.
- [21] H. Bay, T. Tuytelaars, and L. V. Gool, “Surf: Speeded up robust features,” in *ECCV*, 2006, pp. 404–417.
- [22] C. Harris and M. Stephens, “A combined corner and edge detector,” in *Proceedings of the 4th Alvey Vision Conference*, 1988, pp. 147–151.
- [23] T. Lindeberg, “Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention,” vol. 11, 1993, pp. 283–318.
- [24] —, “Edge detection and ridge detection with automatic scale selection,” vol. 30, 1996, pp. 465–470.
- [25] D. Nistr and H. Stewnius, “Scalable recognition with a vocabulary tree,” in *CVPR*, 2006, pp. 2161–2168.
- [26] J. Sivic and A. Zisserman, “Video Google: A text retrieval approach to object matching in videos,” in *Proceedings of the International Conference on Computer Vision*, vol. 2, Oct. 2003, pp. 1470–1477. [Online]. Available: <http://www.robots.ox.ac.uk/~vgg>
- [27] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, “Object retrieval with large vocabularies and fast spatial matching,” in *CVPR*, 2007.
- [28] —, “Lost in quantization: Improving particular object retrieval in large scale image databases,” in *CVPR*, 2008.
- [29] J. L. Bentley, “Multidimensional binary search trees used for associative searching,” vol. 18, no. 9. New York, NY, USA: ACM, Sep. 1975, pp. 509–517. [Online]. Available: <http://doi.acm.org/10.1145/361002.361007>
- [30] B. Goethals and M. J. Zaki, Eds., *FIMI ’03, Frequent Itemset Mining Implementations, Proceedings of the ICDM 2003 Workshop on Frequent Itemset Mining Implementations, 19 December 2003, Melbourne, Florida, USA*, ser. CEUR Workshop Proceedings, vol. 90. CEUR-WS.org, 2003.
- [31] S. Tsang, Y. Koh, and G. Dobbie, “RP-Tree: rare pattern tree mining,” 2011, pp. 277–288.
- [32] B. T. Mark J. Huiskes and M. S. Lew, “New trends and ideas in visual concept detection: The mir flickr retrieval evaluation initiative,” in *MIR ’10: Proceedings of the 2010 ACM International Conference on Multimedia Information Retrieval*. New York, NY, USA: ACM, 2010, pp. 527–536.
- [33] C. Borgelt, “Efficient implementations of apriori and eclat,” in *Proc. 1st IEEE ICDM Workshop on Frequent Item Set Mining Implementations (FIMI 2003, Melbourne, FL)*. CEUR Workshop Proceedings 90, 2003, p. 90.



(a) Stripes and parallel lines



(b) Dots and checks



(c) Single lines



(d) Bright dots

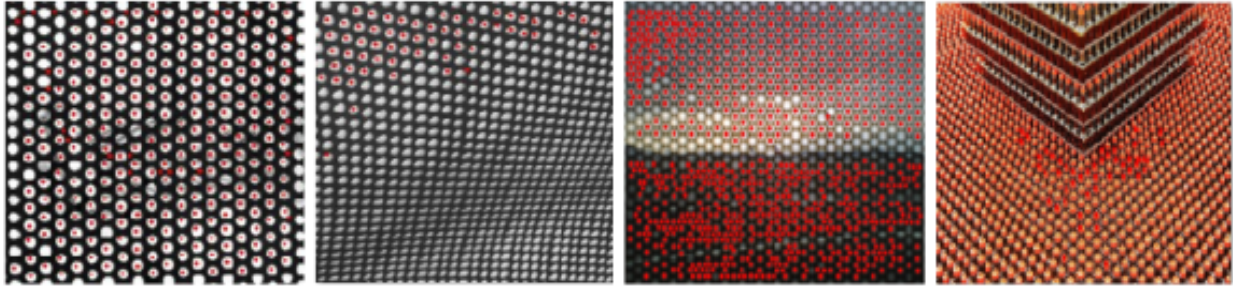


(e) Intersections



(f) Frames

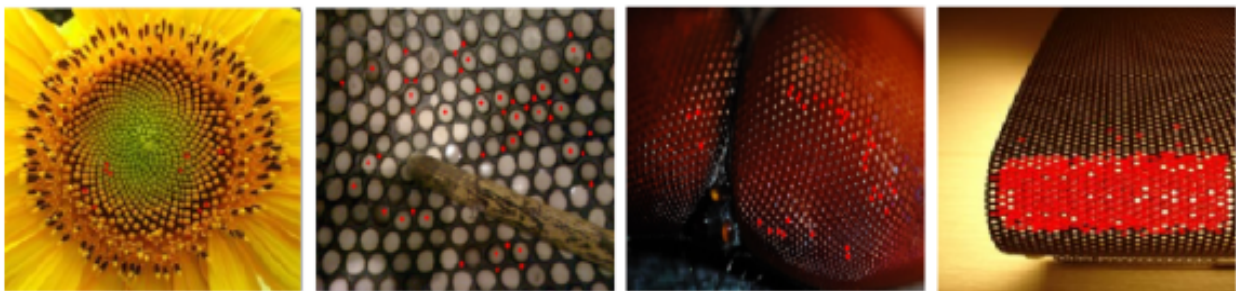
Figure 2. Itemsets representing different patterns in the images. The elements of the itemset are SIFT keypoint identifiers whose location is specified by a red dot placed on the image.



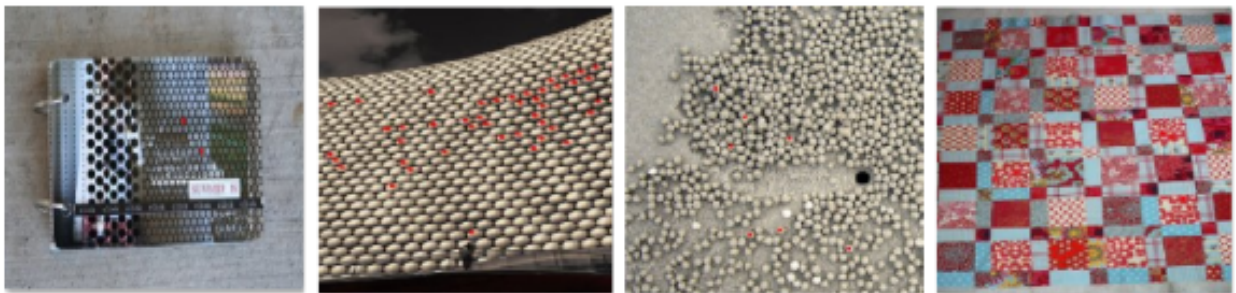
(a) Images containing 4 rare items



(b) Images containing 3 rare items



(c) Images containing 2 rare items



(d) Images containing 1 rare item

Figure 3. Dots and checks: The only semantic pattern observed by rare itemsets mining