SCALABLE SPATIAL EVENT REPRESENTATION

Jelena Tešić, Shawn Newsam, and B. S. Manjunath

Electrical and Computer Engineering Department
University of California, Santa Barbara
Santa Barbara, CA 93106-9560
{jelena, snewsam, manj}@ece.ucsb.edu

ABSTRACT

This work introduces a conceptual representation for complex spatial arrangements of image features in large multimedia datasets. A novel data structure, termed the Spatial Event Cube (SEC), is formed from the co-occurrence matrices of perceptually classified features with respect to specific spatial relationships. A visual thesaurus constructed using supervised and unsupervised learning techniques is used to label the image features. SECs can be used to not only visualize the dominant spatial arrangements of feature classes but also discover non-obvious configurations. SECs also provide the framework for high-level data mining techniques such as using the Generalized Association Rule approach. Experimental results are provided for a large dataset of aerial images.

1. INTRODUCTION

As technology advances and more visual data are available, we need more effective systems to handle the image data processing and understanding. The framework must efficiently summarize information contained in the image data; it must provide scalability with respect to the nature, size and dimension of a dataset; and it must offer simple representations of the results and relationships discovered in the dataset.

On the user side of the system, modelling of a high-level human concept, such as a perceptual event, also raises many research questions. Humans can instantly answer the question "Is this highway going through a desert?" just by looking at an aerial photograph of a region. This query, essentially formulated as a high-level concept, cannot be answered by most existing intelligent image analysis systems. Existing image representations based on low-level features fail to capture perceptual events. Meaningful semantic analysis and knowledge extraction require data representations that are understandable at a conceptual level.

This paper presents an approach to spatial event representation and image analysis at a conceptual level. Section 2 describes the image representation model; Section 3 describes the analysis model, knowledge discovery and analysis techniques; Section 4 presents conducted experiments; and we conclude with a discussion in Section 5.

2. VISUAL THESAURUS

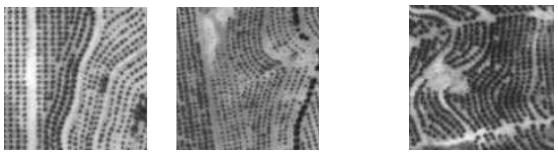
An image analysis framework requires a representation that allows fast data processing, meaningful data summarization, scalability with respect to dataset size and dimension, multi-feature representation, and efficient data understanding. Limited success towards this end has been achieved by systems that use low-level visual features, such as texture and color descriptors, to represent the images. However, these systems fail to support high-level perceptual interaction. A visual thesaurus provides summarized data information derived from the low-level features [1].

2.1. Image Features

The first step in constructing a visual thesaurus is feature extraction. Feature extraction is localized by partitioning an image into tiles. Regular partitioning is a simple alternative to segmentation that allows straight-forward feature extraction and provides a simple spatial layout.

An MPEG-7 [2] compliant homogeneous texture feature vector is extracted for each tile. The 62-dimension feature vector is composed of the first and second order statistics of Gabor filter outputs and Euclidean distance is used the similarity measure. The MPEG-7 homogenous texture descriptor effectively captures visual similarity as shown in the online MPEG-7 demonstration [3]. Other features can be similarly extracted from the tiles.

This research was supported in part by the following grants/awards: The Institute of Scientific Computing Research (ISCR) award under the auspices of the U.S. Department of Energy by the Lawrence Livermore National Laboratory under contract No. W-7405-ENG-48, ONR/ASSERT award #N00014-98-1-0515, NSF Instrumentation #EIA-9986057, and NSF Infrastructure #EIA-0080134.



(a) class 16 (b) class 17

Figure 1: Data Classification Example

2.2. Feature Classification

The second step in constructing a visual thesaurus is feature classification. Conceptually, visually similar tiles are assigned the same class label by partitioning the high dimensional feature space. This is accomplished using a combination of supervised and unsupervised learning techniques.

A set of training tiles is used to configure a Kohonen Self-Organizing Map (SOM). An SOM converts complex, nonlinear statistical relationships between high dimensional data items into simple geometric relationships on a low-dimensional display, while preserving the topological layout of the feature space [4]. The output nodes of the SOM are labelled using the training set and a majority-vote principle [5]. The labels are manually assigned to a training set so that adjacent class numbers correspond to visually similar classes. An example of three training tiles from two agricultural classes is shown in Figure 1.

An initial set of class clusters is formed by using the trained SOM to label each of the dataset features. The Learning Vector Quantization (LVQ3) algorithm is iteratively applied to refine the class clusters [4]. LVQ is a supervised extension of the winner-take-all algorithm [4]. The supervised learning stage of the feature classification is summarized in the following:

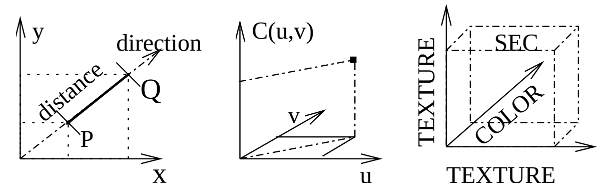
Algorithm 1 Feature Classification

```
SOM summarizes input training feature space;
label SOM output using training set; t = 1.
while (t ≤ T) do
    LVQ fine-tuning of class boundaries;
    re-assign labels using majority-vote approach;
    t = t + 1
end while
```

The unsupervised learning stage of the feature classification further partitions the classes into sets of codewords, as described next.

2.3. Thesaurus Entries

High-dimensional feature spaces are usually very sparse so that enforced space partitioning, such as that described above, frequently clusters visually dissimilar features into the same class. Data partitioning via the Generalized Lloyd Algorithm [6] is used to further split the classes into more consistent clusters. A representative codeword is selected for each cluster and forms the visual thesaurus entry. The remaining cluster features are synonyms of the codeword and receive the same codeword label.



(a) $P\rho Q$ example (b) $C_\rho(u,v)$ (c) Multimodal SEC

Figure 2: Spatial Event Cube

3. SPATIAL EVENT MINING

The motivation for building a spatial event data structure is to discover interesting spatial patterns in extended image datasets. Towards this end, we introduce SECs, a novel data representation obtained by applying spatial predicates to image features labelled using the visual thesaurus.

3.1. Spatial Event Cubes

Spatial Event Cubes are a scalable approach to mining spatial events in large image datasets based on the spatial co-occurrence of perceptually classified image features. Define the image raster space R, for an image partitioned into $M \times N$ tiles, as:

$$R = \{\bar{r} | \bar{r} = (x, y),\ x \in [1, M],\ y \in [1, N]\}$$

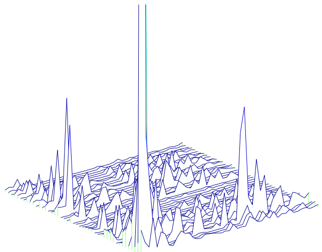
Spatial relationships between coordinates in an image can be defined as a binary relation ρ, $\rho : R \times R \to \{0, 1\}$, or $P\rho Q \in \{0, 1\}$, where $P, Q \in R$. Figure 2(a) shows an example of binary relation ρ, where is ρ defined as a spatial function of distance and direction.

Consider the set of thesaurus entries defined as T, i.e. $T = \{u_i | u_i$ is a thesaurus entry/codeword$\}$. Let τ be the function that maps image coordinates to thesaurus entries, i.e. $\tau : R \to T$, or $\tau(P) = u$, where $P \in R$ and $u \in T$.

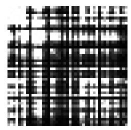
A face of a Spatial Event Cube is the co-occurrence matrix $C_\rho(u, v)$ of thesaurus entries $(u, v) \in T$ of all points whose spatial relationship satisfies ρ:

$$C_\rho(u, v) = \|(P, Q)|\ (P\rho Q) \wedge (\tau(P) = u) \wedge (\tau(Q) = v)\|$$

Figure 2(b) shows the structure of an SEC. $C_\rho(u, v)$ is the number of tiles with thesaurus entries u and v that satisfy spatial relationship ρ. A multi-modal SEC structure is a hypercube whose dimensions are defined by image features extracted from the image tiles. A three-dimensional example, with two texture axes and one color axis, is shown in Figure 2(c).



(a) 3D



(b) 2D

Figure 3: Homogenous Texture Region Analysis in an Image Dataset - SEC Visualization

3.2. Generalized Association Rule

Association rules were introduced as a way of discovering interesting patterns in transactional databases [7]. Frequent item sets are identified using the Apriori algorithm and the most "interesting" rules are selected based on confident factors [8]. Thesaurus entries and their spatial relationships define a non-traditional space for data mining applications. This space can be used to discover interesting rules such as the spatial co-occurrence of orchard and housing regions in aerial images. SECs allow us to extend the traditional association rule approach to multimedia databases.

An attribute value set T contains N thesaurus entries u_i. The SEC entries $C_\rho(u, v)$ mark the frequency of codeword tuples that satisfy binary relation ρ. Define F_K^ρ as a set of frequent item sets of size K. Multiple entry item sets, for $K > 2$, will reduce to ones of smaller order, with different entries. Define $S_\rho^{(K)}$ as a minimum support value for item $(u_1, u_2, ..u_K)$, $(u_1, u_2, ..u_K) \in F_K^\rho$. Our goal is to find $F^\rho = \bigcup_K F_K^\rho$, i.e. sets of tuples that show some dependency among tile spatial configurations. An outline of the extended association rule algorithm for spatial relationships follows:

Algorithm 2 Generalized Association Rule

1. Find frequent item sets;
$F_1^\rho = \{u_i | C_\rho(u_i, u_i) > S_\rho^{(1)}\}$
$F_2^\rho = \{(u_i, u_j) | C_\rho(u_i, u_j) > S_\rho^{(2)}\}$
for $(K = 3;\ F_K^\rho \neq \text{Ø};\ K++)$ **do**
  Candidate K-item frequent itemset is formed of K joint elements from any frequent F_{K-1}^ρ item set;
  Form F_K^ρ from candidates that satisfy the following:
    a. ordering rule of item indices;
    b. minimum support rule;
end for
$F^\rho = \bigcup_K F_K^\rho$.
2. Use the frequent itemsets to generate rules.

The following experiments demonstrate the use of SECs for mining spatial relationships in a large image dataset.

i	22	32	26	35	41
$C_\rho(u_i, u_i)$	24298	20970	18030	8368	7133

Table 1: Codeword Elements of the First-order Item Set F_1^ρ and Corresponding Frequencies

(i, j)	26,2	32,11	22,8	26,46	332,315
$C_\rho(u_i, u_j)$	855	672	633	552	445

Table 2: Codeword Elements of the Second-order Item Set F_2^ρ and Corresponding Frequencies

4. EXPERIMENTS

The proposed visual mining framework is applied to a dataset of 54 large aerial images of the Santa Barbara region. The MPEG-7 homogeneous texture descriptor has shown to be effective at characterizing a variety of land-cover types from this dataset [3]. Each 5248x5248 pixel image is divided into 128x128 pixel non-overlapping tiles resulting in a dataset of 90,744 tiles. A 62-dimension texture feature vector is extracted for each tile.

A visual thesaurus of the tiles is constructed, as described in Section 2. A set of manually labelled tiles is used to train the supervised learning stage of the classification algorithm (Section 2.2). This training set contains 60 land-cover classes, such as agricultural fields, water, parking lots, etc. The 60 classes are further partitioned into 308 codewords using the data clustering techniques described in Section 2.3. These codewords form the thesaurus entries. Every tile in the dataset is labelled with one of these codewords.

SECs are constructed using tile adjacency as the spatial relation. Adjacency is defined as the 8-connectivity neighborhood.

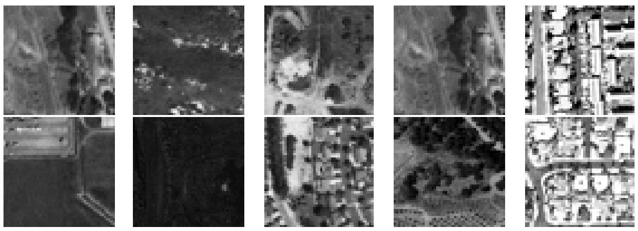
4.1. Visualization

The dominant spatial arrangements of the labelled image tiles over the entire dataset are readily observable from the SEC faces or cross-sections. An SEC faceplate subspace can be visualized as a three-dimensional graph or a two-dimensional image, as shown in Figures 3(a) and 3(b) respectively. The X and Y axes of the graph correspond to classes and the Z axis indicates the relative co-occurrence of two classes with respect to the spatial relation. When an SEC faceplate is viewed as an image, the co-occurrence value corresponds to image intensity.

Figure 3 shows a faceplate of the SEC for the 60 classes in the aerial image dataset using adjacency as the spatial relation. We expect large homogeneous regions in the dataset to result in large values along the diagonal of the faceplate. The spike in Figure 3 corresponds to the ocean class. This makes sense since the aerial images contain large regions of the Pacific Ocean.

22 32 26 35 41

Figure 4: Codeword Tiles Corresponding to the Most Frequent Elements in the First-order Item Set F_1^ρ



(26,2) (32,11) (22,8) (26,46) (332,315)

Figure 5: Codeword Tiles Corresponding to the Most Frequent Elements in the Second-order Item Set F_2^ρ

4.2. Mining

The most frequent first and second order codeword item sets for the aerial image dataset are presented in Tables 1 and 2, respectively. The item sets are computed using the 308 codewords of the visual thesaurus and adjacency as the spatial relation. The most frequent elements of the first order item set F_1^ρ correspond to homogeneous regions. Figure 4 shows the corresponding visual thesaurus codewords, namely pasture and ocean tiles. Higher order item sets provide information about adjacencies between tuples of codewords. Figure 5 shows the visual thesaurus codewords for the most frequent elements of the second order item set F_2^ρ. Figure 6 shows a combination of the the most frequent tuples and triples. Ocean and pasture tiles exhibit composite spatial arrangements.

5. DISCUSSION

This work introduces a novel approach to spatial event representation for large image datasets. Image features are classified using supervised and unsupervised learning techniques. Spatial relationships between the labelled features are summarized using Spatial Event Cubes. SECs are shown to be effective for visualizing non-obvious dataset spatial characteristics such as frequently occurring land-cover arrangements in aerial images. SECs also support the extension of the General Association Rule approach to multimedia databases to identify frequently occurring item sets.

We are using SECs to summarize other sizable datasets, such as a multi-terabyte collection of aerial videos of Amazonia. Future research includes using SECs to construct efficient index structures for multimedia database access.

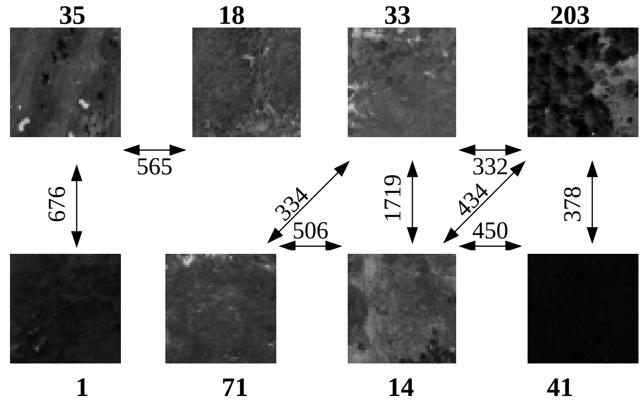


Figure 6: Composite Spatial Arrangement of Ocean and Pasture Tiles in an Aerial Dataset

6. ACKNOWLEDGMENTS

The authors would like to thank Chandrika Kamath and Imola K. Fodor for many fruitful discussions, and Motaz El Saban for extracting the dataset features.

7. REFERENCES

[1] W. Ma and B. S. Manjunath, "A texture thesaurus for browsing large aerial photographs," *Journal of the American Society of Information Science*, vol. 49, no. 7, pp. 633–648, September 1998.

[2] B.S.Manjunath, Philippe Salembier, and Thomas Sikora, Eds., *Introduction to MPEG7: Multimedia Content Description Interface*, John Wiley & Sons Ltd., 2002.

[3] S. Newsam, J. Tešić, M. El Saban, and B.S. Manjunath, MPEG-7 Homogeneous Texture Descriptor Demo: *http://vision.ece.ucsb.edu/texture/mpeg7/*

[4] P. Somervuo and T. Kohonen, "Self-organizing maps and learning vector quantization for feature sequences," *Neural Processing Letters*, vol. 10, no. 2, pp. 151–159, October 1999.

[5] M. Berthold and D.J. Hand, Eds., *Intelligent Data Analysis: An Introduction*, Springer, 1999.

[6] A. Gersho and R.M. Gray, *Signal Coding: Quantization and Compression*, Kluwer Academic Publishers, fourth edition, 1992.

[7] D. J. Hand, H. Mannila, and P. Smyth, *Principles of DataMining*, MIT Press, Cambridge, MA, September 2000.

[8] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in *Proceedings of 20th Internationall Conference on Very Large Data Bases (VLDB)*, September 1994, vol. 3, pp. 487–499.