OBJECT-SPECIFIC FEATURE EXTRACTION VIA MARKOV RANDOM FIELDS DERIVED FROM 0^{TH}-ORDER SIGMA-TREE SEGMENTATIONS

Syed Irteza Ali Khan, *Student IEEE Member*
Christopher F. Barnes, *Senior IEEE Member*
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332, USA
irtezaa@gatech.edu
chris.barnes@gtsav.gatech.edu

ABSTRACT

Sigma-Trees associated with residual vector quantization (RVQ) has been used for image-driven data mining to detect features and objects in a digital image with a degree of success. RVQ methods based on σ-tree structures have been designed to implement successive refinement of information for image segmentation. In such implementations, RVQ based novel methods are devised for pixel-block mining, pattern similarity scoring, class label assignments and attribute mining (Barnes, 2007a). Direct sum σ-tree structures are used for near-neighbor similarity scoring. The variable bit-plane data representations produced by σ-tree structures not only provides an approach for image content segmentation and a structure for formulation of Bayesian classification, but also offers a solution to the challenge of high computational costs involved in pixel-block similarity searching. Such σ-tree based multi-stage RVQ classifiers have yielded promising image-content segmentation/classification yielding fine-grained features extraction. This ability to produce fine-grained features has been exploited in object detection applications. However, in the context of object identification the methods have been applied heuristically on single stages of the multi-stage σ-tree based direct sum successive refinement data representation. As a result, object detection with optimal rejection of false alarm is not guaranteed. Gibbs random field (GRF), also known as Markov random field (MRF), provides a joint probabilistic framework to model the object identification task in digital images. As a result, the image segmentation task can be solved optimally in the maximum aposteriori probabilistic (MAP) sense. Thus, the advantages of the σ-tree based RVQ classifier to provide fine-grained feature extractions for object of interest can be exploited with an MRF-based model of the object. This paper demonstrates the use of MRF on a 0^{th} order output of the σ-tree based RVQ for the purpose of object detection.

INTRODUCTION

Multistage RVQs with optimal direct sum decoder codebooks have been successfully designed and implemented for data compression. The same design concept has yielded good results in the application of image-content classification and has also provided an effective platform to perform image driven data mining (IDDM) (Barnes, 2007a), (Barnes, C. F., Fritz, H., and Jeseon Yoo, 2007b). However, in the context of object identification the methods have been applied heuristically on single stages (0^{th} order segmentation maps) of the multi-stage σ-tree based direct sum successive refinement data representation. As a result, object detection with optimal rejection of false alarm is not guaranteed. GRF provides a practical framework for object detection in images. However, solving the joint probabilistic image model proves to be intractable. The Hammersley-Clifford theorem establishes the Markov-Gibbs equivalence whereby the GRF can be equivalently represented by a Markov random field (MRF). This theorem allows the global property of the GRF to be broken down to an MRF with local property and thus provides an optimal solution in the maximum aposteriori probabilistic (MAP) sense.

The paper is organized in sections. In the following sections the RVQ and MRF models are briefly explained. A section is dedicated to the results of the experiment in which the use of the MRF model on a 0^{th} order segmentation map of the RVQ is demonstrated. The conclusion is drawn in the last section.

RESIDUAL VECTOR QUANTIZATION

Residual vector quantization or quantizer (RVQ), also known as multistage vector quantization or quantizer (MSVQ), have been designed with direct sum codebooks (Barnes, 1993), (Juang and Gray, 1982), (Makhoul, Roucos and Gish, 1985) and (Arnold, 1987). Direct sum codebooks are memory efficient. For example, for an RVQ with P stages and N code-vectors per stage wise codebook, the resultant direct sum codebook contains $\prod_{p=1}^{P} N^p$ code-vectors, but requires memory storage of only $\sum_{p=1}^{P} N^p$ constituent code-vectors, where N^p is the number of code-vectors in the pth stage codebook.

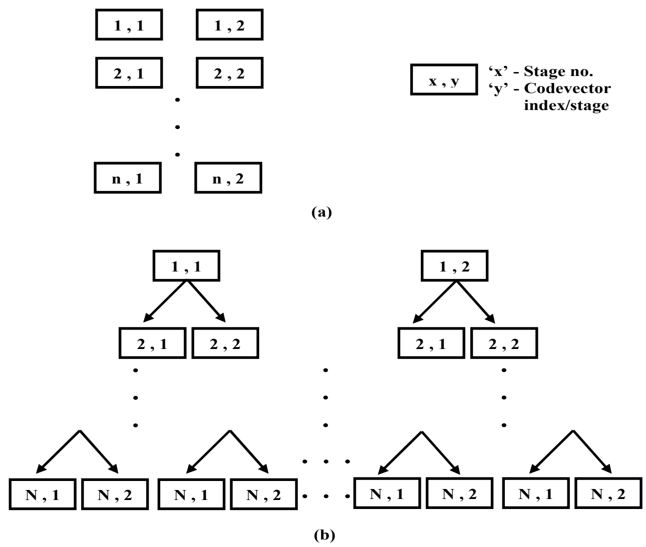


Figure 1. (a) 2-code-vector/stage RVQ with n=N stages or codebooks. **(b)** Direct sum codebook for N-stage RVQ with 2-code-vectors/stage.

Figure 1 illustrates the construction of an RVQ with N stages and two code-vectors per stage wise codebook. The stages are numbered in the top-down manner, where the first stage is the top-most layer and the last stage is the bottom-most layer of the RVQ. Further works focused at improving the direct sum codebook design are (Chan,Gupta and Gresho, 1992), (Barnes,Rizvi and Nasrabadi, 1996). Common to all these design strategies is sub-optimal sequential search encoding, done so to make the RVQ implementation computationally feasible.

After each stage of the RVQ, a segmentation map of the input image can be generated based on the mapping of the input on the respective codevectors at that stage (Barnes, 2007a). Such a segmentation map is termed as 0^{th} order segmentation map. In this paper, object-specific features extraction is performed by imposing MRF framework on the 0^{th} order segmentation map generated from a single stage of the RVQ developed by C. F. Barnes (Barnes, 1993). Figure 2 shows an example of such a segmentation map generated from the second stage of an eight-stage, three-codevector per stage RVQ.

MRF MODEL

Bayesian Labeling Based on MRF

The MRF model used in this paper is based on S. Z. Li (Li, 1995). Let s = {1,2,…,m} be a set of discrete sites and $L^+ = \{0, 1, \ldots, M\}$ be a set of labels which include M physical labels (1,2,….,M) and a virtual $NULL$ label (0). The aim is to assign a label from L^+ to each of the sites in s subject to some contextual constraints. Let $f = \{f_1, f_2, \ldots, f_m\}$

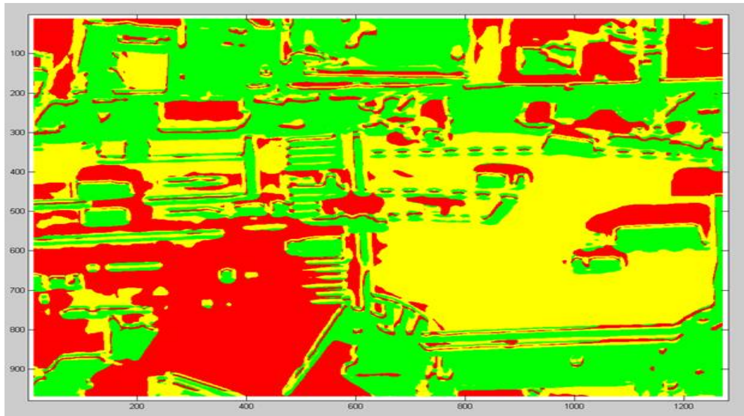


Figure 2. Segmentation map generated from the 2nd stage of 8-stage RVQ with 3-code-vectors/stage.

be a configuration of an MRF with $f_i \in L^+$ assuming a mapping $f: l \rightarrow L^+$ or a labeling of *l*. Let $W = L^+ \times L^+ \times \cdots \times L^+$ (m-time) be the set of all possible configurations.

Given the likelihood function *p*(*r* | *f*) and a priori probability *P*(*f*), the posterior probability can be computed by using the Bayesian rule $P(f|r) \propto p(r|f)P(f)$. The Bayesian labeling problem is the following: given the observation *r*, find the MAP configuration *f** from an admissible space *W*, that is,

$$f^* = argmax_{f \in W}\, P(f|r) \tag{1}$$

According to the Hammersley-Clifford theorem of Markov-Gibbs equivalence (Geman and Geman, 1984), the prior probability *P*(*f*) obeys a gibbs distribution

$$P(f) = Z^{-1} \times e^{-\frac{1}{T}U(f)} \tag{2}$$

Where Z is a normalizing constant, 'T' is a global control parameter called the temperature and *U*(*f*) is the prior energy. The prior energy has the form

$$U(f) = \sum_{c \in C} V_c(f) \tag{3}$$

Where *C* is the set of cliques in a neighborhood system $\mathcal{N} = \{\mathcal{N}_i | \forall_i \in l\}$ for *l* in $\mathcal{N}_i$ is the collection of sites neighboring to *i*.

The likelihood *p*(*r*|*f*) depends how *r* is observed. It can usually be represented in an exponential form $p(r|f) = Z_r^{-1} \times e^{-U(r|f)}$, where $U(r|f)$ is the likelihood energy. Hence the posterior probability is Gibbs distribution $P(f|r) = Z_E^{-1} \times e^{-U(f|r)}$ with posterior energy

$$U(f|r) = U(f)/_T + U(r|f) \tag{4}$$

Therefore, given an observation *r*, a labeling *f* of sites in *l* is also an MRF on *l* with respect to $\mathcal{N}$. The MAP solution is equivalently found by

$$f^* = argmin_{f \in W}\, U(f|r) \tag{5}$$

Posterior Distribution

Neighborhood System and Cliques. In all cases, $\mathcal{N}_i$ can be the set of all the other sites $j \neq i$. This is a trivial case for MRF. In contextual matching, it can consist of all other sites which are related to *i* by the observed realtions in *r*. When the scene is very large, $\mathcal{N}_i$ needs to include only those of the other sites which are within a spatial distance from *i* i.e., $\mathcal{N}_i = \{j \neq i \,|\, dist(feature_j, feature_i) < \alpha, j \in l\}$. The threshold α can be reasonable related to the size of the model object. The set of first order cliques is $C_1 = \{\{i\} \mid i \in I\}$. The set of second order cliques is $C_2 = \{\{i, j\} \mid j \in \mathcal{N}_i, \quad i \in I\}$. In this paper, only cliques of up to order two are considered.

Prior energy. The single site potential is defined as

$$V_1 = \begin{cases} v_{10} & if\ f_i = 0 \\ 0 & otherwise \end{cases}$$

where v_{10} is a constant. This definition implies that a penalty v_{10} is incurred, if f_i is the *NULL* label; or otherwise no penalty. The two sites potential is defined as

$$V_2(f_i, f_j) = \begin{cases} v_{20} & if\ f_i = 0\ or\ f_j = 0 \\ 0 & otherwise \end{cases}$$

where v_{20} is a constant. Similarly, a penalty v_{20} is incurred if either $f_i\ or\ f_j$ is the *NULL*; or otherwise no penalty. The above clique potentials specify the prior energy.

Likelihood Energy. The joint likelihood function $p(r \mid f)$ has the following characteristics: (1) it is conditioned on pure non-NULL matches $f_i \neq 0$, (2) It is independent of the neigborhood system $\mathcal{N}$, and (3) It depends on how the model object is observed in the scene which in turn depends on the underlying tranformations and noise. Assume $r = (r_1, r_2)$ where $r_1 = \{r_1(i) \mid i \in I\}$ and each $r_1(i)$ is a vector of K_1 unary properties; $r_2 = \{r_2(i, j) \mid i, j \in I, i \neq j\}$ and each $r_2(i, j)$ is a vector of K_2 binary relations. The same assumptions are also made for $R = (R_1, R_2)$. The properties and relaitons are assumed to be invariant under the call of underlying transformations for R to r. Assuming an observation model to be $r = R + n$ where n is the independent Gaussian noise, then the likelihood energy is

$$U(r \mid f) = \sum_{i \in I, f_i \neq 0} V_1(r \mid f) + \sum_{i, j \in I, f_i, f_j \neq 0} V_2(r \mid f_i, f_j) \quad (6)$$

because the noise white, we have $U(r \mid f) = U(r \mid f)$ and $U(r \mid f_i, f_j) = U(r_2 \mid f_i, f_j)$. The likelihood potentials are

$$V_1(r_1(i) \mid f_i) = \sum_{k=1}^{K_1} \left[r_{1,k}(i) - R_{1,k}(f_i)\right]^2 \Big/ 2\sigma_{1,k}^2$$

and (7)

$$V_2(r_2(i,j) \mid f_i, f_j) = \sum_{k=1}^{K_2} \left[r_{2,k}(i,j) - R_{2,k}(f_i, f_j)\right]^2 \Big/ 2\sigma_{2,k}^2$$

where $\sigma_{n,k}^2\ (k = 1, 2, \cdots, K\ and\ n = 1, 2)$ are the standard deviations of the noise components. The vectors $R_1(f_i)$ and $R_2(f_i, f_j)$ are the "mean vector" for the random vectors $r_1(i)$ and $r_2(i, j)$, respectively. When the noise is correlated, there are correlatin terms in the likelihood potentials. The assumption of the independent Gaussian noise made may not be accurate but is usually a practical approximation.

Posterior Energy. The posterior energy in (4) can then be derived as

$$U(f \mid I) = \sum_{\{i\} \in C_1, f_i = 0} V_{10}(f_i) \Big/ T + \sum_{\{i,j\} \in C_2, f_i = 0\ or\ f_j = 0} V_{20}(f_i, f_j) \Big/ T + \sum_{\{i\} \in C_1, f_i \neq 0} V_1(r_1(i) \mid f_i) + \sum_{\{i,j\} \in C_2, f_i \neq 0\ or\ f_j \neq 0} V_2(r_2(i,j) \mid f_i, f_j) \quad (8)$$

It can be written into a compact form

$$U(f \mid I) = E(f) = \sum_{i \in I} E_1(f_i) + \sum_{i \in I} \sum_{j \in I} E_2(f_i^1, f_j^1) \quad (9)$$

where

$$E_1(f_i) = \{ V_1(r_1(i)|f_i) \quad \text{otherwise} \}^{\frac{v_{10}}{T}} \quad \text{if } f_i = 0$$

and

$$E_2(f_i, f_j) = \{ V_2(r_2(i,j)|f_i, f_j) \quad \text{otherwise} \}^{\frac{v_{20}}{T}} \quad \text{if } f_j \neq 0 \ \& \ (f_i = 0 \text{ or } f_j = 0)$$

are local posterior energy of order one and two, respectively.

The involved parameters are the noise variances $\sigma^2_{n,k}$ and the prior penalties v_{n0} , with T = 1 fixed. Only the relative, not absolute, values of $\sigma^2_{n,k}$ and v_{n0} are important because the solution *f** remains the same after the energy E is multiplied by a factor. The v_{n0} in MRF prior potential functions controls the behaviour of the system and can be set as desired. The higher the prior penalties v_{n0}, the fewer features in the scene will be matched to the NULL for the minimal energy solution. For the purpose of this particular RVQ framework, the parameters $\sigma^2_{n,k}$ can be ignored and v_{n0} is chosen arbitrarily.

EXPERIMENT RESULTS

Here the results are presented for an object specific-feature extraction application. In the experiment the 0th order segmentation map is generated at the second stage of the RVQ. The aim of this object detection experiment is to detect vehicles in the input image shown in Figure 3(a). The RVQ has eight stages with four codevectors per stage. The segmentation map is shown in Figure 3(b). The color regions associated with the features to be detected are given as follows:-

- Red : Edges
- Green: Top/roof of a vehicle
- Yellow: Other flat objects (like road, pavement, building rooftops etc.)

The MRF model is first order i.e., only pairwise cliques are considered. ‘Edges’, ‘Car-top’ and ‘Other flat surfaces’ are the three features used in the input image. The complete detail of the MRF model is given as follows :-

- Order: 1st order
- Features and their properties
  - Edge: Properties – Length and area of the ‘Edge’.
    Pairwise relation – Distance and angle between two ‘Edges’.
  - Car-top: Properties – Area as the indicator of its size.
    Pairwise relation – Enclosed by ‘Edges’.
    Adjacent to an ‘Edge’.

Table 1. Vehicle-Specific Feature Values

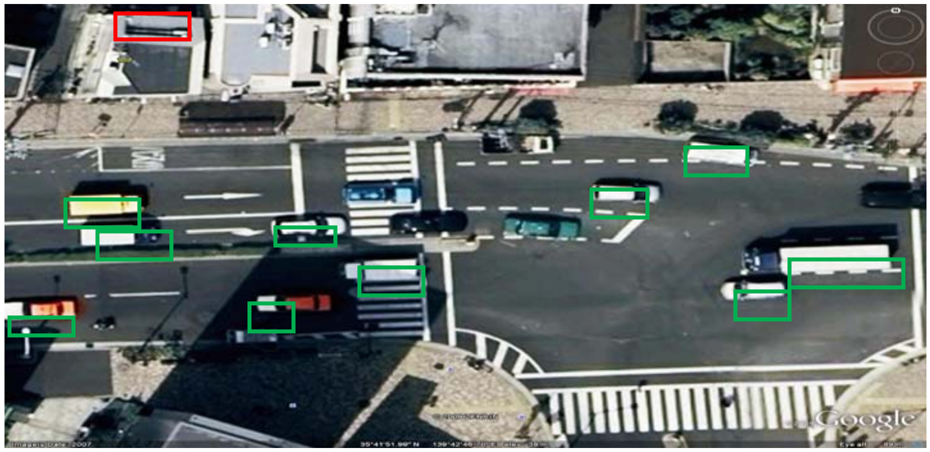
Edge		Car-top	
Length (pixels)	60	Area (pixels)	1200
Area (pixels)	300		
Distance b/w 2 Edges (pixels)	20		
Angle between 2 adjacent Edges (degrees)	90		

The vehicle-specific features are ‘Edge’ (Red region) and ‘Car-top’ (Green region). The pre-specified values of the properties and the pairwise relations between ‘Edge’ and ‘Car-top’ are tabulated in Table 1. The principle for detecting a vehicle using the two features is that these two features in the image must have the properties and relations somewhat similar to the pre-specified properties and relations corresponding to the vehicle object.

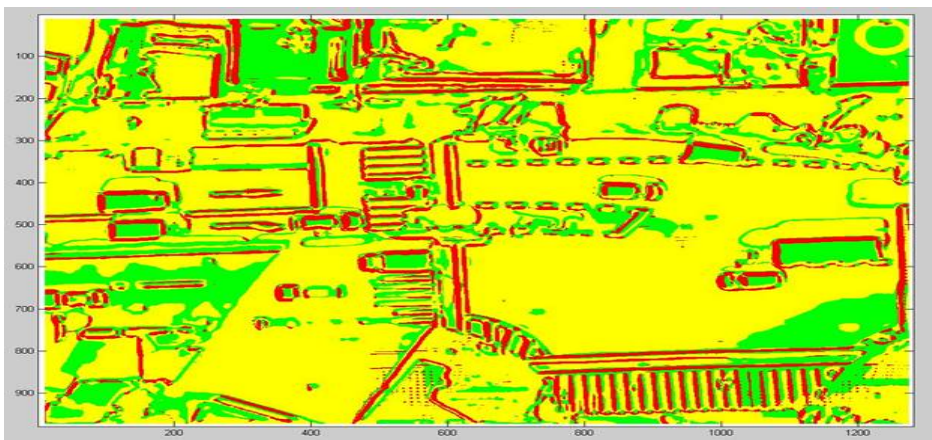
It can be observed in the 0th order segmentation map, figure 3(b), there are considerable false positives for ‘Edge’ and ‘Car-top’. Moreover, in many of the false positives, the two features are also adjacent to each other, thus partially satisfying the pairwise relations of ‘Edge’ and ‘Car-top’. The object detection results of the MRF model is illustrated in figure 2(a). It can be seen that almost all the false positives are rejected. The true-positive detected objects are marked with the green box. The false-positive detections are marked with the red box. The results of the vehicle-detection are tabulated in Table 2.

Table 2. Vehicle Detection Performance

True Positive	False Negative	False Positive
10/15	5/15	1



(a)



(b)

Figure 3. (a) The input image. The true-positive detected vehicles are marked by green box and false-positive by red box **(b)** Segmentation map generated from the 2nd stage of 8-stage RVQ with 3-code-vectors/stage.

CONCLUSION

Markov random field (MRF) provides a suitable framework to model objects based on the object specific features. RVQ, developed by C. F. Barnes (Barnes, 1993), has the capability to generate fine grained features (Barnes, 2007a). The advantages of the σ-tree based RVQ classifier to provide fine-grained feature extractions for object of interest can be exploited with an MRF-based model of the object. With the help of an experiment with simplistic settings, this paper has demonstrated that MRF can be successfully used to correlate object-of-interest specific RVQ features and perform effective detection of the object-of-interest. The MRF-based object detection on a 0th order output of the σ-tree based RVQ is optimal in the MAP sense. However, the efficacy of MRF with the RVQ can only be fully realized when more complicated and comprehensive clique structures, properties and relations are incorporated in the MRF model. Research in this direction is a work in progress.

REFERENCES

Arnold, D. V. 1987. Vector quantization of synthetic array radar data, Master's thesis, Brigham Young Univ., Provo, UT.

Barnes, C. F., 2007a. Image-driven data mining for image content Segmentation, classification, and attribution, *IEEE Trans. Geoscience and Remote Sensing,* vol. 45, issue 1, pp. 2964 – 2978.

Barnes, C. F., Fritz, H., Jeseon Yoo, 2007b. Hurricane disaster assessments with image-driven data mining in high-resolution satellite imagery, *IEEE Trans. Geoscience and Remote Sensing,* vol. 45, issue 6, part 1, pp. 1631 - 1640.

Barnes, C. F. 1993., Vector quantizers with direct sum codebooks, *IEEE Trans. Info. Theory,* vol. 39, issue 2, pp. 565 – 580.

Barnes, C. F., Rizvi, S. A. and Nasrabadi, N. M., 1996. Advances in residual vector quantization: A review," *IEEE Trans. Image Proc*, vol. 5, issue 2, pp. 226-262.

Chan, W.Y., Gupta, S. and Gresho, A., 1992. Enhanced multistage vector quantization by joint codebook design, *IEEE Trans. on Comm*, vol. 40, No. 11, pp. 1693-1697.

Geman, G. and Geman, D., 1984. Stochastic relaxation, gibbs distribution and bayesian restoration of images, *IEEE Trans. Paterrn Analysis and Mach. Intelligence*, PAMI (6): 721-741.

Juang, B. H. and Gray, A. H., 1982. Multiple stage vector quantization for speech coding, *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, vol. 7, pp. 597-600.

Li, S. Z., 1995. A Markov random field model for object matching under contextual constraints," *Proc. IEEE CVPR'94*, pp. 866-869.

Makhoul, J., Roucos, S. and Gish, H., 1985. Vector quantization in speech coding, *Proc. IEEE*, vol. 73, issue 11, pp.1551-1588.