

Detection of Building in Natural Images with one New Discriminative Random Fields

Yanchang Xiao^{1, 2}, Qing Wang¹ and Xiaoguo Zhang¹

*1 School of Instrument Science and Engineering, Southeast University,
Nanjing, China*

*2 School of Information and Engineering, China Jiliang University,
Hangzhou, China
zxcg519@sina.com*

Abstract

This paper presents a new Discriminative Random Fields (DRFs) framework. Based on the DRFs framework proposed by Kumar and Hebert, the following improvements have been conducted. Firstly, the interaction potential and the associated potential model are simplified. Secondly, we reduce the dimension of the multi-scale features, re-definedimension of the single-scale feature, and increase the color feature of Building. Thirdly, the quasi-Newton method with linear search and gradient descent method are adopted to solve parameters, which get a simple model and achieve good performance. Finally, the partition function of the DRF is eliminated by using Pseudo-likelihood method for parameter learning. The simulation results show that the proposed method's false positive rate is lower than the method from Kumar and Hebert, while the correct rate and detection rate are higher than their experimental effects after these improvements.

Keywords: *Building detection; DRF; image classification; quasi-newton method; pseudo-likelihood method*

1. Introduction

Building detection in natural scene is a very active research in some fields such as computer vision, pattern recognition, machine learning, and image processing. Probabilistic methods have been applied to the problem with varying degrees of success, which are Markov random fields (MRFs) [1], conditional random fields (CRFs) and discriminative random field (DRFs).

MRF builds model with the interaction of image elements (pixels or areas), which is used to process image and is applied widely [2-4]. In 2001, Lafferty et al. proposed a classic Conditional Random Field (CRF) model [5], which is the probabilistic models evolved from the MRF. Essentially, CRF is a MRF model under the conditions of the given observation set. CRF is put forward to improve the Maximum Entropy Markov Model (MEMM) and solve the label bias problem based on the directed graph model. Under normal circumstances, the CRF can get better effective than the MRF image processing effects since avoid improperly modeling brings deviation due to the use of the observation field of global information. CRF directly on the posterior probability is modeled as a MRF instead of modeling on the prior probability and likelihood function individually.

CRF was primarily applied to one-dimensional signal processing, such as voice data processing. In 2003, Kumar and Hebert [6] made CRF apply to two-dimensional signal processing and created the DRF framework. It is the primary difference between DRF and CRF that CRF is one-dimensional while DRF is two-dimensional, that creates a loop graph which

has given rise great difficulties to the parameter learning and reasoning calculation. Association potential and interaction potential all adopted the local discriminative classifier, which can take advantage of the interrelated knowledge from special field to make appropriate design, and not rigidly adhere to a fixed form [7].

This paper will make several improvements in the DRFs framework proposed by Kumar and Hebert. Which include simplifying the interaction potential and the associated potential model, changing the features extraction strategies and parameter learning methods. The simulation results show that the proposed algorithm's undetected rate and false detection rate are lower than the method from Kumar and Hebert, and detection effect is also better than their experimental effects after these improvements.

2. Image Model

The description of an image will follow the notation and work of Kumar and Hebert [6]. Images are composed of sites (not necessarily individual pixels), and the classification of an image consists of determining the correct labels of each site in an image. Letting x_i denote the label of the i image site, then $x_i \in \{-1, 1\}$, indicating a site is natural or man-made, respectively. Observed data from an image site i is represented by y_i , and observed data is generated from the feature vectors of the image sites.

Before proceeding to the model of an image the following definition a CRF is given (taken directly from [6]).

Definition of a CRF: Let $G=(S, E)$ be a graph such that x is indexed by the vertices of G . Then (x, y) is said to be a conditional random field if when conditioned on y , the random variables x obey the Markov property with respect to the graph: $p(x_i | y, x_{S-\{i\}}) = p(x_i | y, x_N)$, where $S-\{i\}$ is the set of all nodes in G except the node i , N is the set of neighbors of the node i in G , and Ω represents the set of labels of the nodes in set Ω .

When modeling an image with a DRF, the vertex set corresponds to the set of image sites, and the edge set corresponds to the connections between neighboring sites. In Kumar and Hebert's DRF model for images, they use the Hammersley-Clifford theorem [8] and the assumption that only pairwise clique potentials are non-zero, that is, only immediate neighbors interact. From this they obtain a joint distribution over the labels given observations y which can be written as,

$$P(x|y) = \frac{1}{Z} \exp \left\{ \sum_{i \in S} A_i(x_i, y) + \sum_{i \in S} \sum_{j \in N_i} I_{ij}(x_i, x_j, y) \right\} \quad (1)$$

With a slight abuse of notation, in the rest of the paper we will call A_i the association potential and I_{ij} the interaction potential. There has only relation with a single variable x_i , the A_i associates with a pair of adjacent variables x_i and x_j , the I_{ij} is the neighborhood of the variable x_i , but they are associated with the observed quantity y not just. That note that the DRF can model with the rich features. Here we call Z the partition function, which is actually the sum of all y 's values. The Z can be defined as,

$$Z = \sum_x \exp \left\{ \sum_{i \in S} A_i(x_i, y) + \sum_{i \in S} \sum_{j \in N_i} I_{ij}(x_i, x_j, y) \right\} \quad (2)$$

Here Z is a normalizing factor, it plays mainly standardized role so that the $p(x|y)$ satisfies the conditions that the sum of the probability is one. That reflects the characteristics of the DRF modeling the $p(x|y)$ directly. The x is fixed in the reasoning process, the joint probability distribution $p(x|y)$, established by this generative model like the HMM need to enumerate all

values of x so that it can't reasoning due to the complex model. Furthermore, and may contain a variety of useful characteristics, which includes the complex dependencies on observed data as well as between observation data and the unknown data. It can be seen that for the DRF can directly make conditional probability model to simplify the reasoning calculation, at the same time to express complex features. But the normativity for the sum of probability to be one no longer exists after directly using these complex characteristics, when the y takes over the various values, the potential function can't equal to 1, and so the DRF need to be normalized. However, it brings difficulties to the DRF that the partition function needs to enumerate all y 's values. This is the cost for the DRF to get some advantages.

2.1. Association Potential

In the DRF framework, the association potential $A(x_i)$ is modeled with a local discriminative model that it establishes the relationship of the site with class. Generalized Linear Models (GLM) are used extensively in statistics to model the class posteriors given the observations [9]. For each site i , let $f_i(y)$ be a function that maps the observations y on a feature vector such that $y \rightarrow f_i(y)$. Using the logistic function as the link, the local class posterior can be modeled as,

$$P(x_i|y) = \frac{1}{1 + e^{-x_i(\omega_0 + \omega_1^T f_i(y))}} = \sigma(x_i(\omega_0 + \omega_1^T f_i(y))) \quad (3)$$

Where $\omega = (\omega_0, \omega_1)$ are the model parameters. To extend the logistic model to induce a nonlinear decision boundary in the feature space, a transformed feature vector at each site i is defined as,

$$h_i(y) = [1, \phi_1(f_i(y)), \dots, \phi_R(f_i(y))]^T \quad (4)$$

Where $\phi_k(\cdot)$ are arbitrary nonlinear functions. The first element of the transformed vector is kept as one to accommodate the bias parameter ω_0 . Further, since $x_i \in \{-1, 1\}$, the probability in Eq. (3) can be compactly expressed as,

$$A_i(x_i, y) = \log(P(x_i|y)) = \log(\sigma(x_i \omega^T h_i(y))) \quad (5)$$

Where $f_i(y)$ is a multi-scale feature, which is acting on the three scales. Each scale c ($c = 1, 2, 3$) is corresponding to a histogram site s_c . To get a measure of interaction between sites and their neighbours, multi-scale features were extracted from orientation histograms of block sizes 16×16 , 32×32 and 64×64 , about the center of each image site. Due to the information on the three scales, $f_i(y)$ has drawn two characteristics: intra-scale features and inter-scale features.

2.2. Interaction Potential

The associated potential judge only from local features without global view so that you easily draw fragmented labels, and some blocks away from the building is improperly identified as building. However, the buildings are organic combine complete objects usually, so the labels should gather distribution, the interaction potential is a constraint identification distribution from the overall, which is defined as,

$$I(x_i, x_j, y) = \beta \left\{ K x_i x_j + (1 - K) \left(2 \sigma(x_i x_j v^T u_{ij}(y)) - 1 \right) \right\} \quad (6)$$

Where β is the weight as balance between the association potential A and interaction potential I . Includes two parts, one part is called I sing model, which is independent of the observed data, penalizes every dissimilar pair of labels by the cost and supports the similar labels together. And the other part is $\sigma(x_i x_j v^T u_{ij}(y))$, which is similar to association

potential, and is discriminant model about a pair of labels. It depends on the observed data and is used to smooth adjustment of the extreme constraint I sing model.

we are interested in learning a pairwise discriminative model $p(x_i = x_j | \psi_i(y), \psi_j(y))$ where $\psi: y \rightarrow \mathcal{I}$. Note that by choosing the function ψ to be different from ψ used in Eq.(3), information different from ψ can be used to model the relations between pairs of sites. By extracting different features, ψ will obtain different information with, However, Kumar and Hebert adopted simply the completely equivalent feature of ψ and ψ . Actually they spliced together ψ_i (with ψ_j) to form long $2l +$ vector, the first term is 1 as well as ψ_j . Let $u_{ij}(\psi_i(y), \psi_j(y))$ be a new feature vector such that $u_{ij}: R^r \times R^r \rightarrow R$ denoting this feature vector as u_{ij} for simplification. v and u are all parameter vectors, thus, $2\sigma(x_i x_j v^T u_{ij}(y))$ will map $\sigma(x_i x_j v^T u_{ij}(y))$ to the range of $[-1, 1]$, i.e. is in the same range of x_i , where $0 \leq K \leq 1$, which gives the flexibility to the model by allowing the learning algorithm to adjust the relative contributions of these two terms according to the training data.

2.3. Simply Model

The experimental results from Kumar and Hebert note that the optimal value of K is 0.83, which indicates that between the I sing model independent of observed data and the other part dependent on the observed data exists redundancy. We has repeated Kumar's experiment, which confirmed that the impact of K on the objective function extremum is minimal. Further, it notes that Eq.(6) unnecessarily adopt two parts in the same time. So that, we deleted the Ising items, because there is no reason to impose a smoothing effect regardless of the observed and deny any discontinuity. Moreover, we simplified the kept term, so the interaction potential is modeled as,

$$I(x_i, x_j, y) = x_i x_j v^T u_{ij}(y) \quad (7)$$

In the same time, the association potential is simplified as,

$$A_i(x_i, y) = x_i \omega^T h_i(y) \quad (8)$$

So, the DRF model is defined as,

$$P(x|y) = \frac{1}{Z} \exp \left\{ \sum_{i \in S} x_i \omega^T h_i(y) + \sum_{i \in S} \sum_{j \in N_i} x_i x_j v^T u_{ij}(y) \right\} \quad (9)$$

3. Feature Extraction

Kumar and Hebert used the multi-scale characteristics of 14 dimensions, and reached up to 119 dimensions after the quadratic transformation, with the addition of 28 dimensions characteristics of interaction potential, totally there were characteristics of 147 dimensions. But the parameters has 151(149 plus 2) dimensions, which caused great difficulties for parameter learning. Moreover, if it was not optimized, the three scales calculation is equivalent to 21 times of the size of the original block. Based on these considerations, it has been simplified in this paper, using a single-scale features i.e. 16 16 blocks inter-scale feature. Such considerations also because that the interaction potential includes the characteristic information

of the domain blocks, so it was considered redundant that the associated potential contained multi-scale information.

This paper defines a single scale 10-dimensional feature. The characteristics in the scale of the former four dimensions are substantially similar to the original paper [6]. In addition, it also contains the strongest gradient orientation, density, shape consistency, color monotonicity, the primary colors, and secondary colors. First, the paper cuts the value of the partial derivative matrix obtained, that is, the absolute values of the partial derivative both in the x-direction and y-direction are smaller than a certain threshold value are truncated to 0, therefore leaving only the strong gradient portion, which forms a sparse matrix. Denseness refers to the gradient non-zero number of pixels after processed within the range of four neighbor fields, then making average value within the block is the denseness, and its variance is the shape consistency. So the continuous lines have large dense values and the scattered multi-point gap is smaller; the simple lines will have smaller shape consistency value while the values of messy graphics are bigger. The original paper did not use the color feature, but in fact the color feature is very important for the recognition. This paper statistics the color histogram, the color monotonicity refers to the number of the histogram sites whose value is zero divided by the number of straight square column which is greater than the average, and write down the two primary colors which is used to the interaction potential.

Original features = (Gradient mean, gradient variance, right angel character, verticality, main gradient orientation, denseness, shape consistency, color monotonicity, primary colors, secondary colors)T

In the original paper[6], the associated potential features is equivalent to the interaction potential features f. The f is defined as, f=(Gradient mean, gradient variance, right angel character, verticality, denseness, shape consistency, color monotonicity)T is defined as,

=(Gradient mean, gradient variance, main gradient direction, denseness, shape consistency, color monotonicity, primary colors, secondary colors)T. If do not perform a non-linear transformation, there is

$h = (1, u = (1, |\psi_i \cdot T) T$, the last-dimensional feature of u

is not simply color subtraction, but take the smallest color difference, $\min \{ | \text{primary color} - \text{secondary color} |, | \text{primary color} - \text{secondary color} |, | \text{secondary color} - \text{primary color} |, | \text{secondary color} - \text{secondary color} | \}$. So the u has 8 dimensions instead of 9 dimensions.

The f occasionally has 8 dimensions too. When performing a nonlinear transformation, h and u are simultaneously doing quadratic transformation, but in the original paper [6], u didn't do the non-linear transformation as the dimension of h was high enough. In this paper, h and u are made non-linear transformation to get a total of only 90 dimensions.

At last, we normalized each dimensional feature extracted, by the formula,

$$f_i^m = \frac{f_i^m - \min\{f_i^m\}}{\max\{f_i^m\} - \min\{f_i^m\}} \quad (10)$$

Where the i is the i -dimensional feature, the m is the number of image; normalization is made in an image. Normalization is helpful to parameter learning and the parameters learned are also comparable, that is proven in the analysis of experimental results.

4. Parameter Learning

This paper conducts supervised learning through m artificial identified images by pseudo-likelihood, but the objective function is the logarithmic likelihood,

$$L(w, v) = \sum_{m=1}^M \sum_{i \in S} \log \left(\sigma \left(x_i^m w^T h_i(y^m) + \sum_{j \in N} x_i^m x_j^m v^T u_{ij}(y^m) \right) \right) \quad (11)$$

The parameter to be solved by,

$$(\hat{w}, \hat{v}) = \max_{w, v} \arg L(w, v) \quad (12)$$

This article solves parameters by a quasi-Newton method (BFGS) with line search and gradient descent algorithm. Parameter partial derivatives,

$$\frac{\partial L}{\partial w} = \sum_{m=1}^M \sum_{i \in S} \sigma \left(-q(x_i^m, x_{N_i}^m, y^m) \right) x_i^m h_i(y^m) \quad (13)$$

$$\frac{\partial L}{\partial v} = \sum_{m=1}^M \sum_{i \in S} \sum_{j \in N} \sigma \left(-q(x_i^m, x_{N_i}^m, y^m) \right) x_i^m x_j^m u_{ij}(y^m) \quad (14)$$

The form of the above partial derivatives is simpler than the original paper, and the second derivative is less than zero (with the exception of the mixed second derivative), the experiments results shows that all local maxima of the objective function are global maxima.

5. Eliminate Partition Function

Since using the pseudo-likelihood to carry the parameter learning, we can completely eliminate the partition function in this case by the logistic function. Finally, the modeling may be defined as,

$$P(x_i | x_{N_i}, y) = \frac{1}{1 + e^{-q(x_i, x_{N_i}, y)}} = \sigma \left(q(x_i, x_{N_i}, y) \right) \quad (15)$$

$$q(x_i, x_{N_i}, y) = x_i \omega^T h_i(y) + \sum_{j \in N} x_i x_j v^T u_{ij}(y) \quad (16)$$

Because the value of $q(x_i, x_{N_i}, y)$ may not be in the range of $[0, 1]$, that does not comply with the probability definition. So it may be mapped to the legitimate range with the logistic function, this avoids the problems of partition function. Now, we will prove that the model without partition function is equivalent to it with the partition function.

The process of demonstration is as,

If $Q(x_i) = A(x_i, y) + \sum_{j \in N_i} I(x_i, x_j)$, then

$$\begin{aligned}
 P'(x_i|x_{N_i}, y, \theta) &= \frac{1}{z_i} \exp\{Q(x_i)\} \\
 &= \frac{\exp\{Q(x_i)\}}{\exp\{Q(x_i)\} + \exp\{Q(-x_i)\}} \\
 &= \frac{1}{1 + \exp\{Q(-x_i) + Q(x_i)\}}
 \end{aligned}$$

And

$$Q(-x_i) - Q(x_i) = 2 \left\{ x_i \omega^T h_i(y) + \sum_{j \in N} x_i x_j v^T u_{ij}(y) \right\},$$

if there is $\begin{cases} \omega = 2, \\ v = 2, \end{cases}$ then the two is equivalent. That's over.

6. Experiments and Discussion

6.1. Input Image and Feature Calculation

The training and the test set all contained 81 images, each of size 256 384 pixels, all from the Corel image database. Each image was divided in non-overlapping 16 16 pixels blocks, and we call each such block an image site. The ground truth was generated by hand-labeling every site in each image as a structured or non-structured block. The whole training set contained 31,104 blocks from the non-structured class and from the structured class respectively. The experimental procedure is as follows, firstly, input training image and label data, after reading the labeled graph, they were converted, -1 presents the non-building and 1 presents building, that is consistent with the above. Secondly, calculating the original features based on the above context's methods, we selected 16 angles of the gradient histogram and selected 10 colors from color histogram. Thirdly, calculating the value of h and u. Fourthly, if you wanted to make the non-linear conversion you would make quadratic mapping. Fifthly, inference and sampling by Gibbs. Finally, calculating correct rate, false positive and detection rate. The above process took about 1.5 seconds per image averagely, while the process took more than a dozen seconds with faithful to the original feature extraction methods.

6.2. Experiment Results

6.2.1. Feature Extraction Results: Feature Extraction has a great effect on detection results, so we developed a visualization experiment platform to observe meticulously of the feature extraction process. As shown in Figure 1, the current process is the ninth and the fifteenth blocks of the first picture, the white box deviation to right of central in the upper two pictures indicates the current block, and it is moving along with the experimental image processing block. The RGB picture shows the original image, the upper left image is grayscale picture for feature extraction. From left to right in the lower is the gradient histogram, color histogram, the gradient intensity of the current block and grayscale respectively. The bottom shows some of the features extracted, which are respectively the gradient mean, gradient variance, right angle character, verticality, denseness, shape consistency, color monotonicity from left to right.

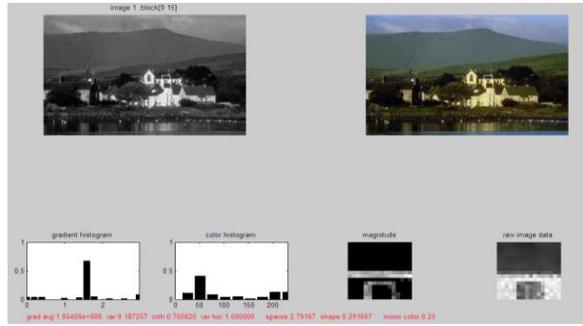


Figure 1. Feature Extraction Experiment Platform

Before feature extraction, the partial derivative matrix was processed in the directions of x and y. The absolute values which are less than 40% of the strongest partial derivative were all set to 0, so as to project the intensity gradient and form a sparse partial derivative (gradient) matrix. The value of 40% is empirical value which set up according to observations. Processing results is shown in figure 2. By the top two figures (a) and (b), it is shown that most of the information is preserved after the cut-off value, and the two images have not a little differences. But the vast majority of the details are all excluded, as we can see from the figure (c) and figure (d). The gradient matrix vast majority is non-zero value before processing in the figure (c) which was all white, but after processed, the vast majority becomes zero, leaving a clear outline of the major information. The bottom two figures (e) and figure (f) are similar to the above, Gradient direction also becomes zero after processed. The compare of computing performance and results is shown in Table 1, which all adopted the quadratic nonlinear mapping features and quasi-Newton method. After cutoff, the time of calculating value saved 2/3 and the value of logarithmic likelihood was 0.55% less than before.

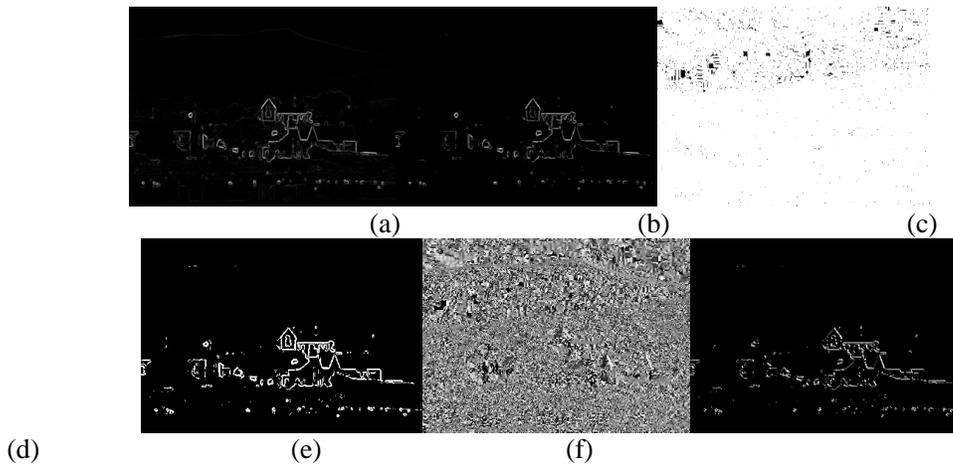


Figure 2. Gradient intercept contrast, the left column of the picture shows the original situation, the right column is the situation after the process of cut-off; the top pictures of (a) and (b) are grayscale pictures which mapped to the rang of [0, 255]; pictures (c) and (d) are binary image, that grayscale gradient nonzero is 255, otherwise, gray is 0. The bottom pictures of (e) and (f) are gradient direction graph. The direction of the $[0, \pi]$ corresponding to the gradient histogram subscript $[1, 16]$, the subscript value representing the direction is mapped to the rang of [0,255].

Now, we analysis what the various features played a role. For analysis, we did not make a non-linear mapping to features. Because of the features carried out normalization processing, the symbols and the value of the parameters learned could reflect effect of feature. Parameter results are shown in Table 2 (obtained by the gradient descent method).

Table 1.Gradient Cut-off and Results Contrast

	The mean time (each image)	The value of pseudo-likelihood
Notcut-off	5s	-2159.47
Cut-off	1.4s	-2171.42

Because the reference numerals of the building is a positive number 1 and all the features are non-negative, so w_i is the positive value, which shows that w_i play a positive role in discriminant buildings (the bigger of feature value, the more likely it is the building), otherwise, at the opposite side (the smaller of the feature values, the more likely it is a building), so as to make the logistic function and logarithmic likelihood to be greater. The table 2 shows that the first five parameters of w are positive (bias constants are not features), the parameter of shape conformability is negative, this is coincide with the original designed DRF. Only color monotonic is contrary with the original intention. We thought that the building's color is single; instead, the building's color in the natural scenery is mottled. But the results from Figure 2 show that the more mottled color of building is more like a building.

In addition, the greater of the absolute value of the parameters, the better of degree of distinction, and the discrimination of gradient mean, gradient variance, verticality and color monotonicity are very much, and right angles character, denseness and shape consistency are not meaningful. About v , the gradient mean difference and the color difference played a leading role, especially the difference of color is up to 1.15, that illustrates v playing a good role on discriminating the pair of labels consistency, and this feature was not used in [6].

Table 2. Parameter Learning Results

Parameters	w	v
Bias constants	-0.37446	-1.6367
Gradient mean	0.73043	0.96767
Gradient variance	0.39499	0.1386
Right angle	0.16329	0.13081
Verticality	0.48225	-0.26965
Denseness	0.043644	0.28458
Shape consistency	-0.15547	-0.1808
Color monotonicity	-0.78158	1.1522

6.2.2. Parameter Learning Method Contrast: The quasi-Newton method BFGS was comprehensively superior to other method from Table 3.

However, it has been observed that the gradient descent method in the initial stage of parameters learning close to the optimum value is faster than the quasi-Newton method; it is harder to close to the optimal value until the later period.

Table 3. Parameter Learning Method contrast

Methods	Quasi-Newton	Gradient Descent
Iteration times	35	51
Linear searchmean times	4	6.5

Logarithm Pseudo-likelihood	-2264.6	-2292.83
Gradient vector norm	8.78	18.1
Time-consuming	456.41s	1006.4s

6.2.3. Detection Results: The standards to measure the results of good or bad are all different, there is no a standard to fully reflect the problem. There are three amounts to represent the detection results of structure are as following,

Correct rate (CR), it is correct even if be equal to the label number. Each image's correct number is divided by the label total number (16 24), and then count mean of all images.

False positive (FP), it is the probability of which the non-building is mistakenly regarded as a building.

Detection rate (DR), it is the probability of which the building is correctly identified as the building. The number of blocks in the area of the building whose label is correct divided by the total number of buildings blocks.

As shown in the table 4, the non-linear features did not bring about substantial improvement, which stated that characteristics defined in this paper have a good linear judged scope without quadratic nonlinear mapping. In addition, parameter learning methods do not significantly affect the pros and cons of the test results. P mean is the mean value of the formula (14) which calculated by pseudo-likelihood, to some extent, it reflects the correct rate of learning parameters on the judgment of the training images. The difference between the P mean and correct rate might explain shortcomings of the Gibbs algorithm so that reasoning algorithm can be enhanced in the future. The time-consuming of Gibbs algorithm is about 0.4s per image equally.

Table 4. Detect Results

Feature	Parameter learning method	CR	FP	DR	P(Mean)
Linear	Quasi-Newton	86.7%	10.8%	66.1%	92.978 %
	Gradient Descent	85.3%	13.1%	72.5%	92.894 %
QNL	Quasi-Newton	86.0%	10.5%	59.9%	93.257 %
	Gradient Descent	88.2%	8.4%	66.1%	93.004 %

As shown in the Figure 4, the detection result of the improved DRF proposed in this paper compared to the detection results from Kumar's experiments. The "real situation", refers to the manually annotated results, are from <http://www.cs.cmu.edu/~skumar/>, which we painted on the box to the corresponding position on the original image. For this sample image, the detection rate (marks on the house) from this paper is better than Kumar and Hebert [6], however the false detection rate (marks on the tree) is too bad. Tim Rees [10] of UBC repeated the experiment of Kumar and Hebert [6] too, the detection results from this paper was compared with his results in Figure 5.

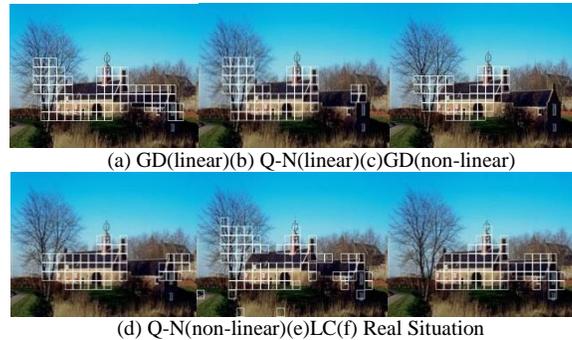


Figure 4. Detection Results of One Image Selected with the Above Methods



Figure 5. Results of the Two Images compared with Tim Rees [10]

The left is result of Tim Rees, the middle is the results of this paper, and the right is a real situation.

7. Conclusion

In this paper, we present an improved Discriminative Random Fields (DRFs). Based on the DRFs framework proposed by Kumar et al, we do the following modification. We simplified the interaction potential and the associated potential model, reduced the dimension of the multi-scale features, imports the color feature of building and adopted the quasi-Newton method with linear search and gradient descent method to solve parameters. The simulation results show that the detection results are better than Kumar experimental effects. The next step is to extend the model to accommodate multiclass classification problems. In the future, we also should seek better building features (such as texture features etc.) for image classification.

Acknowledgments

This work is supported by the National Key Technologies R&D Program of China during the 12th Five-year Period (2013BAJ13B01).

References

- [1] P. Perez, "Markov random fields and images", *CW Quarterly*, vol. 11, (1998).
- [2] U. Kurkure, Y. H. Le, N. Paragios, T. Ju, J. P. Carson and I. A. Kakadiaris, "Markov RandomField-based fitting of a subdivision-based geometric atlas", *International Conference on Computer Vision*, Barcelona, Spain, (2011), pp. 2540-2547.

- [3] L. Cordero-Grande, G.Vegas-Sanchez-Ferrero, P. Casaseca-de-Ia-Higuera and C. Alberola-Lopez, "A Markov Random Field approach for Topology-Preserving Registration: application to Object-Based Tomographic Image Interpolation", IEEE Transactions on Image Processing, vol. 21, no. 4, (2011), pp. 2047-2061.
- [4] J. Sun and M. F. Tappen, "Learning non-local range Markov Random field for image restoration, Computer Vision and Pattern Recognition, Colorado Springs, CO, USA, (2011), pp. 2745-2752.
- [5] J. Lafferty, A. McCallum and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data", In Proc. ICML, (2001).
- [6] S. Kumar and M. Hebert, "Discriminative random fields: A discriminative framework for contextual interaction in classification", In IEEE International Conference on Computer Vision (ICCV), Nice, France, (2003).
- [7] C. Sutton and A. McCallum, "An Introduction to Conditional Random Fields for Relational Learning", (2007).
- [8] S. Z. Li, "Markov Random Field Modeling in Image Analysis", Springer-Verlag, Tokyo, (2001).
- [9] P. McCullagh and J. A. Nelder, "Generalised Linear Models. Chapman and Hall, London, (1987).
- [10] T. Rees, "Detection of Man-made Structures in Natural Images", (2004).

Authors



Yanchang Xiao, he received respectively his B.S. degree in the Department of Computer Science and M.S. degree in Applied Mathematics from Chengdu University of Technology, China. In 2004, he started the teaching career in School of Information and Engineering, China Jiliang University. Since 2010, he studied in the Department of Instrument Science and Technology from Southeast University as a doctoral candidate. His research interests are in the areas of image processing, pattern recognition, machine learning.



Qing Wang, he was born in 1962 in China. He got his Ph.D. degree in the Department of Instrument Science and Technology from Southeast University, China, in 1998. Now he is a professor in the Department of Instrument Science and Engineering, Southeast University, China. His current research field is GPS/GIS theory and application.



Xiaoguo Zhang, he is now with the School of Instrument Science and Engineering, Southeast University, as an associate professor. He received the M.S. and Ph.D. degrees in the CAD and GPS/DR/MM integrated vehicle navigation field from Southeast University, Nanjing, China, in 1998 and 2001, respectively. His current research fields are focused on image processing, GIS and 3S-integration.