A Multilayer Markovian Model for Change Detection in Aerial Image Pairs with Large Time Differences

Praveer Singh, Zoltan Kato, Josiane Zerubia

A Multilayer Markovian Model for Change Detection in Aerial Image Pairs with Large Time Differences

Praveer Singh
Telecom ParisTech
46 rue Barrault
F-75634 Paris Cedex 13, France
Email: praveer.singh@telecom-paristech.fr

Zoltan Kato
Institute of Informatics
University of Szeged, 6701 Szeged,
P. O. Box 652, Hungary
Email: kato@inf.u-szeged.hu

Josiane Zerubia
INRIA, AYIN team,
2004 route des Lucioles,
06902 Sophia Antipolis, France
Email: josiane.zerubia@inria.fr

***Abstract*—In this paper, we propose a *Multilayer Markovian model* for change detection in registered aerial image pairs with large time differences. A *Three Layer Markov Random Field* takes into account information from two different sets of features namely the Modified HOG (Histogram of Oriented Gradients) difference and the Gray-Level (GL) Difference. The third layer is the resultant combination of the two layers. Thus we integrate both the texture level as well as the pixel level information to generate the final result. The proposed model uses pairwise interaction retaining the sub-modularity condition for energy. Hence a global energy optimization can be achieved using a standard min-cut/ max flow algorithm ensuring homogeneity in the connected regions.**

I. Introduction

Change Detection is of utmost interest in the field of remote sensing be it for tree / boat detection, forest fire, urban planning or disaster management in hazardous areas [1] [2] [3]. Various state of the art techniques have been introduced in the previous years with the advent of stochastic modeling in image processing. One of the first work in this domain [4], laid down Markov Random Fields (MRF) as an effective model for image restoration. However the energy minimization problem is NP-hard which requires powerful optimization tools like Simulated Annealing [5], or sub-optimal deterministic techniques such as ICM [6], Modified Metropolis Algorithm [7], or in some special cases of energy functions [8] the famous graph based min-cut/max flow algorithm [9].

In case of complex segmentation problems, a single layer based MRF model is not sufficient to obtain the desired results. *Multilayer Markov Random Field* (MMRF) have successfully been used in the earlier works of [10] [11] [12] [13] as well. [10] has been able to successfully detect changes in aerial images with small time differences using a three layer Markovian model. [12] [13] were able to detect changes with large time differences in aerial image pairs but this time by employing a Mixed Markov field [14].

In the proposed approach, we have tried to include both the texture and pixel level information to build a three layer Markov model using a *Modified Histogram of Oriented Gradients (HOG)* [15] and the *Gray Level (GL) Difference* features on the topmost and bottommost layer respectively. Using the precomputed Ground Truth (GT) mask (done manually by an expert), we employ a supervised technique to mark the initial set of pixels / sites as foreground or background. With the *Modified HOG difference* [15] and the *Gray level difference* feature vector computed for all the pixels in the image pairs, a probability density function is fitted individually for the foreground and background labels. This probabilistic estimate is calculated using one training image pair for each of the 3 data sets (same data set as used in [13]) . Using this probabilistic measure of the background and foreground labels, a negative log likelihood is computed for each pixel corresponding to the 2 feature vectors. This is then fed as input to the energy function of the proposed 3-layer MRF model. The final segmentation is obtained through the optimization of sub-modular energy via graph-cut [9], as a result of which we obtain foreground and background labeling over the combined layer (the third layer in between the Modified HOG and GL difference layer in the 3-layer MRF model). The proposed model is validated over all the different image pairs present in the three data sets obtained from the Hungarian Institute of Geodesy, Cartography and Remote Sensing using the manually generated ground truth [13].

In the next section, we describe the image model and the feature selection in detail. Multi-layer MRF (MMRF) segmentation model is discussed in Section III. MAP estimate via graph-cut is illustrated in Section IV. Experiments are described in Section V. Finally Conclusion is given in Section VI.

II. Image Model and Feature Selection

Feature selection is carried out over the aerial image pair data set (see again [13]). The registered image pairs I_1 and I_2 are over the same region but with a large time difference and with different atmospheric conditions. Image pairs were converted to gray scale-images $\widetilde{G_1}$ and $\widetilde{G_2}$ and then normalized by taking the L_2-norm over all the image pixels given by:

$$G_k = \frac{1}{\sqrt{\left(\sum_{i,j} \left(G(i,j)\right)^2\right)}} * \widetilde{G}_k$$

This work has been partially funded by INRIA and by the European Union and the State of Hungary, co-financed by the European Social Fund through project TAMOP-4.2.4.A/2-11-1-2012-0001 National Excellence Program and through project FuturICT.hu (grant no.: TAMOP-4.2.2.C-11/1/KONV-2012-0013).

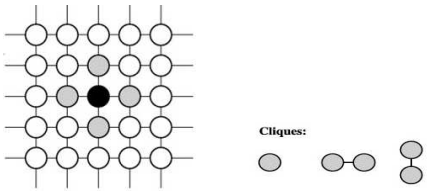


Fig. 1: Image Model. (a) Lattice S with a 1st order neighborhood system. (b) All possible cliques for S.

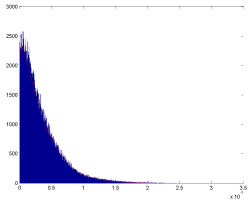


Fig. 2: Histogram corresponding to Background pixels for the GL difference feature fitted with a Generalized Gamma distribution.

where k = {1,2}. To improve the results further and get a better quality of the features we employ a smart contrast normalization technique (see [16]).

A. Image Model

The two image pairs are built over the same lattice S which is composed of sites (or pixels) s such that $S = \{s_1, s_2..., s_N\}$ as described in Figure 1. The two registered gray scale images G_1 and G_2 have their gray scale values respectively denoted by $g_1(s)$ and $g_2(s)$ for a given pixel s∈S. A 4-neighborhood system (figure 1(a)) is defined on the lattice as :

$$\forall s \in S : \Phi_s = \{r \in S : \|s - r\| = 1\}, \tag{1}$$

Hence the cliques can be defined as in figure 1(b). The primary goal is to classify each site s ∈ S as changed (*foreground*) or unchanged (*background*). Hence the assignment of a label to a particular site is from the set : $\lambda = \{fg, bg\}$ where $\{fg\}$ refers to foreground class and $\{bg\}$ refers to the background class. $\{fg, bg\}$ labels will be represented by $\{1, 0\}$ as there respective logical values.

B. Feature Extraction

In the proposed method we try to incorporate two kinds of features, namely the *Gray Level (GL) difference* and the *Modified Histogram of Oriented Gradients difference* (HOG) [15] between the image pairs. We consider the background/ foreground classes as random processes generating the above mentioned features. Hence we try to model these random processes by fitting a suitable distribution function over the histograms (for both features) corresponding to each of the foreground and background class using a set of training image pairs. The training image pairs contain a Ground Truth having all the pixels manually labeled as background / foreground by an expert.

1) GL difference feature: The Gray Level difference feature computed over the image pairs for each corresponding pixel is defined as :

$$d(s) = \|g_1(s) - g_2(s)\| \tag{2}$$

Using the already available ground truth of the training image pair, we plot histogram for the different $d(.)$ values corresponding to both background and foreground pixels. Pixels belonging to background class are estimated by a generalized gamma distribution (as shown in Figure 2) , using the WAFO toolbox [17], whose density is given as :

$$P\left(d(s)|bg\right) = f\left(d(s)|a, b, c\right) = \frac{c}{b^{ac}\Gamma(a)} d(s)^{ac-1} e^{-\left(\frac{d(s)}{b}\right)^c} \tag{3}$$

Since the foreground can have any d(s) value, we try to approximate the foreground class using a uniform density function given as :

$$P\left(d(s)|fg\right) = \begin{cases} \frac{1}{b_d - a_d}, & \text{if } d(s) \in [a_d, b_d]. \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

2) Modified HOG difference feature: The Histogram of Oriented Gradient (or simply HOG) feature that we use is a somewhat modified form of the original method as proposed in [15]. HOG is a feature descriptor that has mainly been used for object detection in the field of image processing. It basically involves counting the number of occurrences of different orientations of gradients inside a bounding box fixed by us and then rounding it to the correct bin of the histogram. In [15] the implementation of the HOG has been achieved by dividing the image into blocks and then further into small cells for which the histogram of gradients is computed. Finally the concatenation of all resulting histograms leads to the descriptor for the entire image. However in our case, instead of already framed cells over the entire image, we employ a sliding window of size 11×11 which is then made to slide over the entire image.
The first step in the HOG implementation involves gradient computation at each pixel. The gradients are computed using the following basic filters:

$$I_x = I * [-1 \;\; 0 \;\; 1] \qquad I_y = I * [-1 \;\; 0 \;\; 1]^T$$

Henceforth the amplitude and the orientation can be calculated using the following equations:

$$\|H\| = \sqrt{I_x^2 + I_y^2} \qquad \theta = \arctan\left(\left\|\frac{I_y}{I_x}\right\|\right)$$

It can be seen that the range of θ is now varying between 0 to $\pi/2$. The second step consists of computing the histogram associated with every position of the sliding window over the entire image. The HOG (with *9 bins*) so computed for a particular position of the sliding window is then assigned as a feature vector $\overrightarrow{f_s}$ to the center pixel of the window. Thus finally after applying this technique, we now have a 9 dimensional vector $\overrightarrow{f_s}$ associated with every pixel s∈S of the image.
In [15], in order to get a better performance, the local histograms are contrast-normalized by calculating a measure of the intensity across the block, and using this value to normalize all cells within the block. However in our case we get better

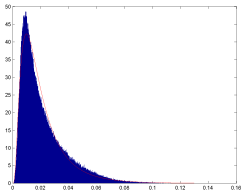


Fig. 3: Histogram corresponding to Background pixels for the Modified HOG difference feature fitted with a Generalized Gamma distribution.

results with an unnormalized HOG because we have already done the contrast stretching as an image preprocessing step as discussed before. We have tried to test various combinations of window and bin sizes, with and without contrast stretching over the entire image as well with and without normalized and unnormalized HOG over the sliding window. The best results were obtained by employing contrast stretching over the entire image, and with an unnormalized HOG computed over a detection window of size 11×11 and using 9 bins. Hence forth the HOG difference corresponding to a particular pixel s is given by:

$$\overrightarrow{h(s)} = \|\overrightarrow{f_{s_1}} - \overrightarrow{f_{s_2}}\| \tag{5}$$

Finally we add up all the terms present in the 9- dimensions of $\overrightarrow{h(s)}$ corresponding to every pixel s in order to obtain a 1-D HOG difference feature h(s) for each site. The background class is approximated using a generalized gamma distribution (as shown in Figure 3) with the WAFO toolbox [17] which is given by:

$$P\left(h(s)|bg\right) = f\left(h(s)|u,v,w\right) = \frac{w}{v^{uw}\Gamma(u)} h(s)^{uw-1} e^{-\left(\frac{h(s)}{v}\right)^{w}} \tag{6}$$

Again the the foreground class is estimated using a uniform density function given by -

$$P\left(h(s)|fg\right) = \begin{cases} \frac{1}{b_h - a_h}, & \text{if } h(s) \in [a_h, b_h]. \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

3) Maximum Likelihood Estimate: The maximum likelihood estimate for the GL difference feature is computed by the following equation:

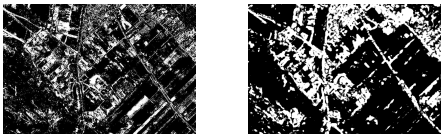
$$\arg \max_{\phi \in \{fg,bg\}} P(d(s)|\phi)$$

Similarly for the Modified HOG difference feature as well, Maximum Likelihood estimate is given by:

$$\arg \max_{\phi \in \{fg,bg\}} P(h(s)|\phi)$$

where $\phi \in \{fg, bg\}$ is the binary label set from which a label is assigned to any given pixel s.

Figure 4(a) and Figure 4(b) show the maximum likelihood results using the GL and the Modified HOG difference features respectively. It is clearly visible that the weak segmentation obtained from the Maximum Likelihood estimate of the two features is not sufficient to classify the pixels. Lot of noise is noticed both in the foreground and the background regions due to the out-lier pixels.



(a) (b)

Fig. 4: Maximum Likelihood Estimate for. (a) GL Difference Feature. (b) Modified HOG Difference Feature.

Henceforth there is clearly a need for a more sophisticated technique, which takes into account the weak segmentation coming from the two features without neglecting prior knowledge on the homogeneity of the regions.

III. Multi-Layer MRF Segmentation Model

As shown in the Figure 5, our proposed multi-layer MRF segmentation model [10] [11] [12] [13] is built over a Graph $\mathcal{G}$ composed of three different layers namely S^h, S^c and S^g all being of the same size as the lattice S of the image model. Each pixel $s \in S$ has a corresponding site associated with it in each of the three layer given as:

$$s^h \in S^h, \qquad s^c \in S^c, \qquad s^g \in S^g$$

where S^h, S^c and S^g correspond to the three different layers of the multi-layer MRF segmentation model representing Modified HOG difference feature, final desired change map and GL difference feature respectively. Every site s^i where $i \in \{h, c, g\}$ has also a class label specified to it denoted by $\omega(s^i)$ which is modeled as a discrete random variable taking values from the label set $\lambda = \{fg, bg\}$. The initial labeling of the sites $s^i \in \{S^h, S^g\}$ is characterized by the weak segmentation results discussed in the previous section for each of the two Modified HOG difference feature and the GL difference feature. All the different sites s^i corresponding to the third layer S^c would be the final desired change map. Thus we can define the hidden label process as the set of all the labels over the entire graph $\mathcal{G}$ as follows:

$$\underline{\omega} = \left\{\omega(s^i) | s \in S, i \in \{h, c, g\}\right\} \tag{8}$$

Since we are dealing with Multi-Layer Markov Random Fields, we defined two kinds of neighborhood system in the proposed technique as clearly shown in the figure 5. The first one corresponds to the intra-layer connections between the site pairs at every layer of the graph $\mathcal{G}$ where we follow the 1st order neighborhood system as discussed in Section II-A. The second one is the inter-layer interaction between the sites of different layers. In other word it consists of 5 interactions between two layers: Any site s^i interacts with the corresponding site s^j (where $\forall s \in S; \{i, j\} \in \{\{h, c\}, \{c, g\}\}$) on the neighboring layer as well as with the 4 neighboring sites two steps further away as shown in the figure 5. The first kind of interactions yields homogeneity within each layer whereas the second kind of interaction ensures that the final change mask coming from the S^c layer does not neglect the information from the other two layers.

Hence the graph $\mathcal{G}$ has doubleton "intra-layer" and "inter-layer" cliques (denoted by $\mathcal{C}_2$ and $\mathcal{C}_5$ respectively). The singleton cliques are denoted by the set $\mathcal{C}_1$ corresponding to the

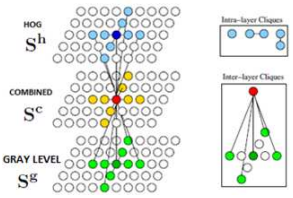


Fig. 5: MRF Segmentation Model. The interlayer and the intra-layer interactions are clearly visible in the Hierarchical model. Also shown are intra layer singletons as well intra and inter layer doubleton's.

TABLE I: Quantitative Results over the 3 data sets. All the False Alarm, the Missed Alarm and the Overall Error in percentage of the number of processed image pixels. We can see that the Overall Error for our model (HMRF) is at least 0.7% less than the CXM model.

Data-Sets	False-Alarm		Missed-Alarm		Overall-Error	
	CXM	HMF	CXM	HMF	CXM	HMF
Szada	2.86	1.10	1.31	2.32	4.18	**3.43**
Tiszadob	2.91	0.06	1.77	3.89	4.68	**3.96**
Archieve	7.59	2.22	3.06	6.42	10.66	**8.65**

one element sets with single individual sites linking our model to the two observation features namely the GL Difference and the HOG difference. Thus the set of cliques is given by:

$$\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2 \cup \mathcal{C}_5 \tag{9}$$

Both the GL difference and the Modified HOG difference features, are represented by the observation process :

$$\mathcal{F} = \{f(s)|s \in S\}, \tag{10}$$

where $f(s) = \{h(s), d(s)\}$.

This noisy image process is basically a deviation from the underlying label process.
Our goal is to find the optimal labeling $\hat{\underline{\omega}}$ which maximizes the a posteriori probability $P(\underline{\omega}|\mathcal{F})$, which is the maximum a posteriori (MAP) estimate [4] given as :

$$\hat{\underline{\omega}} = \arg\max_{\underline{\omega}\in\Omega} P(\underline{\omega}|\mathcal{F}) \tag{11}$$

where Ω denotes the set of all the possible labellings. Based on the Hammersley-Clifford Theorem, the posterior probability for a particular labeling follows a Gibbs distribution:

$$P(\underline{\omega}|\mathcal{F}) = \frac{exp(-U(\underline{\omega}))}{Z} = \frac{1}{Z} exp\left(-\sum_{C\in\mathcal{C}} V_C(\underline{\omega}_C)\right) \tag{12}$$

where $U(\underline{\omega})$ is the energy function, V_C denotes the *clique potential* of $C \in \mathcal{C}$ having the label configuration $\underline{\omega}_C$, Z is a normalization constant independent of $\underline{\omega}$ given by $Z = \sum_{\underline{\omega}\in\Omega} exp(-U(\underline{\omega}))$.
Any clique refers to a set of sites, which may contain a single element or more in such a way that any two distinct elements present in the set are the corresponding neighbors of each other defined by the 1st neighborhood system. Hence the doubleton "intra-layer" cliques corresponding to the sites s^h and r^h can be denoted by $\{s^h, r^h\}$. Similarly for the "inter-layer" cliques for sites s^h and r^c we have the corresponding clique as $\{s^h, r^c\}$. As discussed earlier the singleton potentials are the ones which represent the direct influence of the observation process on the probabilistic modeling of a label of a site without taking into account any contextual information. whereas the doubleton clique potentials do express a strong relationship between the neighborhood pixels.
Since the labellings for S^h and S^g layers are directly influenced by the values of $h(.)$ and $d(.)$ respectively, we have: $\forall s \in S$, the singleton potentials given as:

$$V_{\{s^h\}}(\omega(s^h)) = -\log P(h(s)|\omega(s^h)). \tag{13}$$

$$V_{\{s^g\}}(\omega(s^g)) = -\log P(d(s)|\omega(s^g)). \tag{14}$$

where the probabilistic measure, given that the label being a background or a foreground class generate the h(s) or d(s) observations, has already been defined in Section II. The labels corresponding to the sites of the S^c layer have no direct influence by these observations and hence the singleton potential is set to ϕ. For the intra-layer cliques given by $C_2 = \{s^i, r^i\}$ where $C_2 \in \mathcal{C}_2$ and $i \in \{h, c, g\}$, the potential can be defined as follows :

$$V_{C_2} = f\left(\omega(s^i), \omega(r^i)\right) = \begin{cases} 0, & \text{if } \omega(s^i) = \omega(r^i). \\ 2\mathcal{K}^i, & \text{if} \omega(s^i) \neq \omega(r^i). \end{cases} \tag{15}$$

where $i \in \{h, c, g\}$. $\mathcal{K}^i \geq 0$ is a parameter controlling the homogeneity of the regions. As $\mathcal{K}^i$ increases, the resulting regions in the corresponding layer (based on i) become more homogeneous.
Next for the inter-Layer cliques given by $C_5 = \{s^i, r^j\}$ where $C_5 \in \mathcal{C}_5$ and $\{i, j\} \in \{\{h, c\}, \{c, g\}\}$, the corresponding potentials are given by:

$$V_{C_5} = f\left(\omega(s^i), \omega(r^j)\right) =$$

$$\begin{cases} \rho^{hc} * W_r * \|V_{\{s^h\}}(\omega(s^h)) - V_{\{s^h\}}(\omega(r^h))\|, & \text{(a).} \\ \rho^{cg} * W_r * \|V_{\{s^g\}}(\omega(s^g)) - V_{\{s^g\}}(\omega(r^g))\|, & \text{(b).} \end{cases} \tag{16}$$

where (a) and (b) refer to $\{i, j\} = \{h, c\}$ and $\{i, j\} = \{c, g\}$ respectively. Also $V_{\{s^h\}}(\omega(.))$ and $V_{\{s^g\}}(\omega(.))$ are the singleton potentials for the sites s^h and s^g belonging to the layers S^h and S^g respectively dependent on the labeling $\omega(.)$. Both parameters ($\rho^{hc} \geq 0$ and $\rho^{cg} \geq 0$) control the influence of the feature layers(S^h and S^g) to the combined (S^c) layer. W_r is the weight of the interaction. We assign higher weight (W_r = 0.6) to the corresponding site whereas smaller weights (W_r = 0.1 each) to the other 4 neighboring sites.
Hence the optimal MAP labeling $\hat{\underline{\omega}}$, which maximizes $P(\underline{\omega}|\mathcal{F})$ (i.e. minimizes $-\log P(\underline{\omega}|\mathcal{F})$) can be calculated as the minimum energy configuration as:

$$\hat{\underline{\omega}} = \arg\min_{\underline{\omega}\in\Omega} -\sum_{s\in S} \log P(h(s)|\omega(s^h)) -$$

$$\sum_{s\in S} \log P(d(s)|\omega(s^g)) + \sum_{C_2\in\mathcal{C}_2} V_{C_2}(\underline{\omega}_{C_2}) + \sum_{C_5\in\mathcal{C}_5} V_{C_5}(\underline{\omega}_{C_5}) \tag{17}$$

IV. MAP Estimate via Graph-cut

The multi-layer MRF segmentation model as discussed in the section III consists of binary labels namely the foreground and the background class {fg / bg}. Moreover, the Gibbs energy given in equation (17) is composed of singleton and

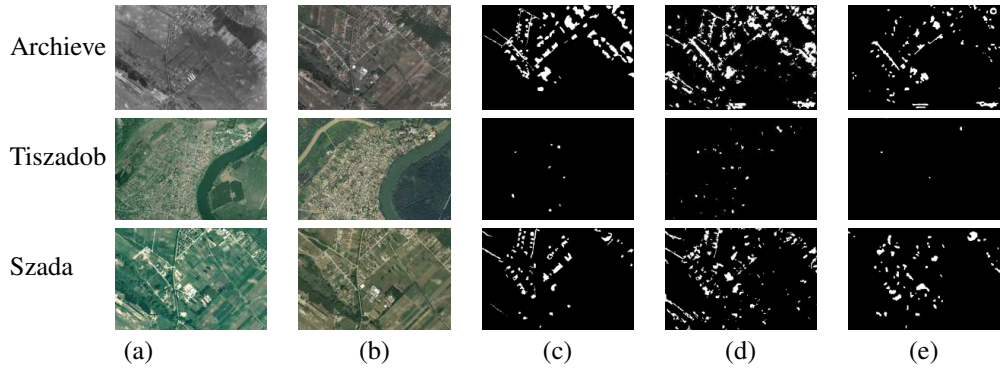


Fig. 6: Qualitative results over 3 data set. Column (a) - image 1; column (b) - image 2; column (c) - ground truth; column (d) - CXM model results; column (e) - our own (MMRF) model results.

doubleton potentials. It is because of the occurrences of the pairwise interactions as well as the binary labels in our model, that the Gibbs energy in equation(17) can be represented in the form of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. The set of vertices's $\mathcal{V}$ in the graph $\mathcal{G}$ consists of all the sites present in the 3-layer MRF segmentation model and two other terminals namely the *source* **s** and the *sink* **t**. There are also edges $\mathcal{E}$ present in the graph $\mathcal{G}$ representing not only the doubleton's but also the connection of all the sites with the two terminals (for more details see [8]). A cut on the graph $\mathcal{G}$ represents a binary partitioning **S**,**T** of the vertices's such that the terminals source $\mathbf{s} \in \mathbf{S}$ and sink $\mathbf{t} \in \mathbf{T}$ and can be described using the binary variable $\omega(s^i)$ where $s \in S$ and $i \in \{h, c, g\}$. The cost of any cut **c(S,T)** is the sum of the edge weights c(u,v) that go from S to T. Hence the energy represented by $\mathcal{G}$ can be seen as a function $E(\underline{\omega})$ equal to cost of the cut defined by $\underline{\omega}$. Thus in order to compute the energy minimization, we need to compute the min cut /max flow in the corresponding graph $\mathcal{G}$. The energy function $E(\underline{\omega})$ can be decomposed as follows :

$$E(\underline{\omega}) = \sum_{s \in S} E_s(\omega(s^i)) + \sum_{(s,r) \in \mathcal{C} - \{\mathcal{C}_1\}} E_{s,r}(\omega(s^i), \omega(r^j)) \tag{18}$$

where i , j $\in \{h, c, g\}$; i = j for intra-layer cliques and inter-layer clique's otherwise. The first term in the above equation refers to the energy corresponding to the singletons whereas the second term corresponds to the doubleton's (both intra-layer and inter-layer). Before applying the graph-cut algorithm, a necessary and sufficient condition for graph-representation of the energy function $E(\underline{\omega})$ is the sub-modularity condition :

$$E_{s,r}(bg, bg) + E_{s,r}(fg, fg) \leq E_{s,r}(bg, fg) + E_{s,r}(fg, bg) \tag{19}$$

which is always true in our case since the left hand side is always 0, while the right hand side is a value always greater than or equal to 0 depending on the singleton potential and the homogeneity constants. Hence we used a standard graph-cut algorithm implemented by Prof. Vladimir Kolmogorov (http://pub.ist.ac.at/ vnk/software.html) to minimize the energy in equation (17).

V. EXPERIMENTAL RESULTS

A. Test Databases

For testing and evaluation, three sets of optical aerial image pairs provided by the Hungarian Institute of Geodesy Cartography and Remote Sensing (FOMI) and Google Earth were used. As described in [13], the data sets namely SZADA, TISZADOB and ARCHIEVE contain 7, 5 and 1 image pairs respectively along with manually generated Ground Truth (GT) by an expert. Also image pairs in the same test set have been taken from nearby regions,with similar camera properties. Data set SZADA contained images taken by FOMI in 2000 and 2005, respectively. Due to limitations of the WAFO toolbox, we used 5 image pairs in this test set, all of them manually evaluated covering in aggregate 9.5 km^2 area at 1.5-m/pixel resolution (the size of each image in the test set is 952 $\times$ 640 pixels). One image pair was used for training and the remaining 4 ones for validation. From the second test set called TISZADOB we used 2 image pairs from 2000 to 2007 (6.8 km^2) with the same size and quality parameters to SZADA. For the last test set ARCHIVE, a comparison was made between the aerial image taken by FOMI in 1984 and a corresponding Google Earth photograph from 2007. This test set was highly challenging due to the degraded quality from the 1984 image pair and several major differences appeared during the 23 year time slot between the two shots. GT masks were generated manually by an expert for each image pair of the data sets.

B. Parameter Estimation

The following parameters are used in the proposed model which need to be estimated :
1. Parameters for Modified HOG feature selection: $\{dw, nb\}$
2. Parameters of the various pdf's as introduced in (3), (4), (6), (7) : $\{a, b, c, u, v, w, a_d, b_d, a_h, b_h\}$
3. Parameters of the intra-layer and inter-layer clique potential functions : $\{\mathcal{K}^i, \rho^{hc}, \rho^{cg}\}$ where $i \in \{h, c, g\}$
The approach used in our model is a supervised one. Hence all the parameters are estimated using the training data provided in each of the three data sets. The initial parameters for the HOG feature selection namely the detection window size dw and the bin size nb were set by evaluating the maximum-likelihood results. By experimentation the desired results were

obtained by setting the detection window to 11×11 and the number of bins to 9.
Parameters for the Generalized Gamma Distribution (in (3), (6)) corresponding to the the background class (both for the gray level and the Modified HOG features) were learned from the training data provided in each of the data sets. The threshold values for the uniform distribution (in (4), (7)) corresponding to the foreground class for both the gray level and the HOG features were set to the optimal values again for each of the training data belonging to its respective data set.
Parameters of the intra-layer and inter-layer clique potential functions were set by trial and error over the training data to their accurate values.

C. Results and Evaluation

In this section we would try to make a comparison of our method with the CXM method (proposed in [13]) which has yielded better results on the same data sets as compared to other state of the art methods as discussed in [13]. The comparisons is made both qualitatively and quantitatively with CXM which also uses a supervised approach as our method.
For a quantitative evaluation, the same metrics is used as in [18] [19]. The segmentation results obtained with the CXM model were compared to the manually generated Ground Truth and henceforth the number of false alarms (unchanged pixels falsely detected as changes), missed alarms (changed pixels which were erroneously ignored during the segmentation) and overall error (total sum of the false alarms and the missed alarms) were computed.
The evaluation rates measured over the 3 data sets are given in table I(overall error) in percentage of the number of processed image pixels. It can be clearly seen that the overall error of our proposed method is at least 0.7% powerful than the CXM model. As stated in [13] the weaker results in the Archieve data sets are mainly due to the poor image quality.
Qualitative results also showed comparative results over some image portions (see figure 6). We can notice that our approach produces more homogeneous and smoothening effect than the CXM model. Using graph cut our method is computationally very efficient (takes 10 -15 seconds / image pairs) compared to CXM method in [13].

VI. Conclusion

This paper proposed a novel approach for change detection in aerial images with large time difference using a hierarchical MRF segmentation model incorporating feature selection in terms of Modified HOG and GL difference and using the min-cut/max-flow algorithm. The approach has outperformed the previous best results from CXM model quantitatively whereas qualitatively the results are quite comparable. The proposed method has taken into account both the pixel level GL difference as well as texture level information in terms of the Modified HOG feature. The hierarchical MRF model takes into account some prior knowledge about the homogeneity of the regions within a layer as well as with the neighboring feature layer which is another advantage. Finally the energy minimization is performed using the graph cut technique which is time efficient and reliable.
Hence the proposed method can be used for various remote sensing applications including assisting operators in evaluation of data sets.

References

[1] P. Gamba, F. Dell' Acqua, G. Lisini, and G. Trianni, "Improved vhr urban area mapping exploiting object boundaries," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 8, pp. 2676–2682, 2007.

[2] A. Farag, R. Mohamed, and A. El-Baz, "A unified framework for map estimation in remote sensing image segmentation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 7, pp. 1617–1634, 2005.

[3] H. Deng and D. Clausi, "Unsupervised segmentation of synthetic aperture radar sea ice imagery using a novel markov random field model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 528–538, 2005.

[4] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, pp. 721–741, 1984.

[5] Z. Kato and J. Zerubia, *Markov Random Fields in Image Segmentation. Collection Foundation and Trends in Signal Processing*. Now Editor, World Scientific, Sep. 2012. [Online]. Available: http://hal.inria.fr/hal-00737058

[6] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society. Series B*, vol. 48, no. 3, pp. 259–302, 1986.

[7] Z. Kato, J. Zerubia, and M. Berthod, "Satellite image classification using a modified metropolis dynamics," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, 1992, pp. 573–576.

[8] V. Kolmogorov and R. Zabin, "What energy functions can be minimized via graph cuts?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 2, pp. 147–159, 2004.

[9] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1124–1137, 2004.

[10] C. Benedek, T. Sziranyi, Z. Kato, and J. Zerubia, "Detection of object motion regions in aerial image pairs with a multilayer markovian model," *IEEE Transactions on Image Processing*, vol. 18, no. 10, pp. 2303–2315, 2009.

[11] Z. Kato and T. Pong, "A multi-layer MRF model for video object segmentation," in *Proceedings of Asian Conference on Computer Vision*, ser. Lecture Notes in Computer Science, P. J. Narayanan, S. K. Nayar, and H.-Y. Shum, Eds., vol. 3852. Hyderabad, India: Springer, Jan. 2006, pp. 953–962.

[12] C. Benedek and T. Sziranyi, "A mixed markov model for change detection in aerial photos with large time differences," in *International Conference on Pattern Recognition*, 2008.

[13] ——, "Change detection in optical aerial images by a multilayer conditional mixed markov model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 10, pp. 3416–3430, 2009.

[14] A. Fridman, "Mixed markov models," *Proceedings of the National Academy of Sciences*, vol. 100, no. 14, pp. 8092–8096, 2003. [Online]. Available: http://www.pnas.org/content/100/14/8092.abstract

[15] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, 2005, pp. 886–893.

[16] S. Koptenko, "Contrast stretch and normalization," http://www.mathworks.fr/matlabcentral/fileexchange/11429-contrast-stretch-and-normalization, 2006.

[17] P. Brodtkorb, P. Johannesson, G. Lindgren, I. Rychlik, J. Rydén, and E. Sjö, "WAFO - a Matlab toolbox for the analysis of random waves and loads," in *Proc. 10'th Int. Offshore and Polar Eng. Conf., ISOPE, Seattle, USA*, vol. 3, 2000, pp. 343–350.

[18] S. Ghosh, L. Bruzzone, S. Patra, F. Bovolo, and A. Ghosh, "A context-sensitive technique for unsupervised change detection based on hopfield-type neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 3, pp. 778–789, 2007.

[19] L. Bruzzone and D. Prieto, "An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images," *IEEE Transactions on Image Processing*, vol. 11, no. 4, pp. 452–466, 2002.