ISSN : 0974 - 7435
Volume 10 Issue 16
2014
BioTechnology
An Indian Journal
FULL PAPER
BTAIJ, 10(16), 2014 [9237-9245]

Face detection based on conditional random fields

Huachun Yang[1,2]
[1]School of Life Sciences and Technology, Xidian University, Xi'an, Shaanxi, 710071, (CHINA)
[2]Engineering University of Armed Police Force, Xi'an, Shaanxi, 710086, (CHINA)
Email: hbyhc331@aliyun.com

ABSTRACT

To address the local occlusion and pose variation in face detection, face can be looked on as a whole composed of several parts from up to down. First, the face is divided into a number of local regions from which various features are extracted. Each region is identified by a local classifier and is assigned a preliminary part label. A random field is established based on these labels and multiple dependencies between different parts are modeled in a CRF framework. The probability that the test image may be a face is calculated by a trained CRF model. The probability is used as a measure to test the existence of a face. The experiments were carried out on the CMU/MIT dataset. As indicated by the experiment results, the following methods can improve the detection rate and enhance the robustness of face detection in case of occlusion: 1) integrating multiple features and multiple dependencies in CRF framework; 2) dividing the face optimally.

KEYWORDS

Face detection; Conditional random fields; Multiple features; Multiple dependencies.

INTRODUCTION

Face detection is an important research topic in the field of computer vision, which aims to determine existence of a face in a static image and to acquire the face's accurate location and scope. The article[1-3] is the review of related literature. Scale and translation invariance are acquired by sliding a window over an input image at different resolutions. The research on face detection involves two issues: effective search strategy and robust classifier for face and non-face. The former affects the detection speed and the latter determines the detection accuracy which is the key issue discussed in this paper.

Face detection methods can be classified into two types. The first one is based on the idea that a face is an indivisible whole entity[4-5]. The second is part-based[6-11], which can adapt to changes in posture and partial occlusion as compared with the former. Part-based methods mainly solve two problems: 1) extracting features from parts and establishing the mapping relationship between local features and face parts; 2) modeling dependencies among face parts.

The mosaic image method presented by Yang et al.[7] divided the original image evenly into rectangular cells and established gray distribution in face. The test image was filtered from a low resolution to a high resolution according to the distribution. The detection rate is not high for only a kind of gray feature is used in this method.

F.S.Samaria and Nefian et al. presented HMM-based face detection[8-10] and consider a face as composed of five parts: hair, forehead, eyes, nose, mouth. They divided the face into a series of rectangle-blocks. In[8], the gray features were extracted from the block, in[9] the DCT coefficients and in[10] the KLT coefficients. Above all, only one kind of feature was extracted. The mapping relationship between rectangle-blocks and face parts was acquired by multi-Gauss model. Due to diversity of faces, it is difficult to determine the number of Gaussian kernel. In addition, the dependencies between parts were modeled by an one-step transition probability matrix for the Markov assumption. However, this assumption is too strict for face detection. For only such a kind of dependency was modeled, local error can be transmitted to other regions.

Besides, different parts and dependencies have different effects on face detection. The detection method presented by Epshtein and Ullman[11] assigned different weights to different parts according to the maximum mutual information. However, each part's matching is very complex for it's almost a detection problem.

Above all, many researchers have tried to divide the face into a series of parts, which plays an important role in face detection. Based on this idea, we try to improve the accuracy and robustness in three ways: 1) extracting various features from each part and integrating them; 2) integrating multiple dependencies between different parts; 3) different features, different parts and different dependencies are assigned different weights by training.

These ideas are implemented in a CRF[12] framework. The CRF presented by Lafferty et al. is a statistical model for labeling and segmenting sequence data, and is capable of processing independent and interaction features. Recent years see an explosion of interest in the CRF with successful applications including text processing[13-14], bioinformatics[15-16], and computer vision[17-20]. In a CRF framework, different weights are acquired for different features, different parts and different dependencies by training. The probability whether a test image may be a face is calculated by using the trained CRF model.

The remaining of the paper is structured as follows. Section 2 introduces face segmentation and feature extraction method. The CRF framework for face detection is then discussed in detail in Section 3. Section 4 presents the experimental method and the results. Finally, we summarize this paper and explain the future work in section 5.

DIVIDING FACE AND EXTRACTING FEATURES

As illustrated in Figure 1, a face image is divided into a sequence of blocks. Each face image of width W and height H is divided into overlapping blocks of height L and width W. The number of overlaps between consecutive blocks is P.

The number of blocks extracted from each face image is defined as the number of observation vectors T and is calculated by formula $T=(H-L)/(L-P)$.

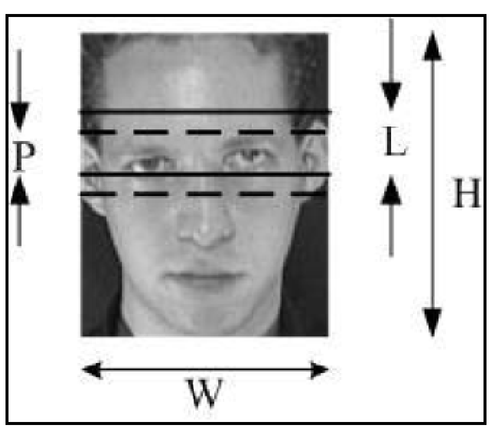


Figure 1: Face image segmentation

The face is considered as a whole composed of five parts (such as hair, forehead, eye, nose, mouth etc.) from top to bottom, and each part corresponds to a part label. Each block which belongs to a part of the face is assigned a part label. T blocks have T labels with mutual dependencies, which constitute a random field in a face. Multiple features are extracted from each block and different types of features have different weights.

CRF-BASED FACE DETECTION

CRF framework

The CRF model is defined as a graphical model used to calculate the probability of a possible class label sequence for a given observation sequence. CRFs use a single exponential distribution to model all labels of the given observations[12]. In our application each class label corresponds to a part in a face and observations correspond to feature vectors extracted from all blocks.

Using the Hammersley Clifford theorem[21] and assuming only up to pairwise clique potentials to be nonzero, the joint distribution over the labels y given the observations x can be denoted as:

$$P_\theta(y \mid x) = \frac{1}{Z_\theta(X)} \exp\left(\sum_{e \in E} \sum_k \lambda_k f_k(e, y \mid e, x) + \sum_{v \in V} \sum_k \mu_k g_k(v, y \mid v, x) \right) \quad (1)$$

$$Z_\theta(X) = \sum_y \exp\left(\sum_{e \in E} \sum_k \lambda_k f_k(e, y \mid e, x) + \sum_{v \in V} \sum_k \mu_k g_k(v, y \mid v, x) \right) \quad (2)$$

Where λ and μ are CRF model parameters, which are a trade-off in the weights of each feature function f and g.

The function g can be interpreted as a measure of the compatibility between an observation x and class label y, g (x, y) is implemented using a local discriminative model that outputs the association of the observation x with class label y. Libsvm[29] is used extensively in statistics to model the class posteriors given the observations and outputs N (N is the number of part label) probabilities that x may be the label yi (i=1,2,…N). In a block, different features such as Discrete Cosine Transform (DCT) and local binary pattern (LBP)[22,30], are extracted. Accordingly, different libsvm models are trained for every feature respectively. Thus, g k(x, y) corresponds to the k th feature. Different kinds of feature have different importance on the result and μ_k denote the weight of the k th kind of feature.

Different scales denoted by w means different intervals and dependencies. In practice, two-scale dependencies are involved: w1=1 and w2=T/5. Two matrixes P1 and P2 are used to express the corresponding part transition.

$$P_1 = \begin{pmatrix} (S-1)/S & 1/S & 0 & 0 & 0 \\ 0 & (S-1)/S & 1/S & 0 & 0 \\ 0 & 0 & (S-1)/S & 1/S & 0 \\ 0 & 0 & 0 & (S-1)/S & 1/S \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$P_2 = \begin{pmatrix} 1/S & (S-2)/S & 1/S & 0 & 0 \\ 0 & 1/S & (S-2)/S & 1/S & 0 \\ 0 & 0 & 1/S & (S-2)/S & 1/S \\ 0 & 0 & 0 & 1/S & (S-1)/S \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

P1 and P2 are determined by the face segmentation method. For w=1 means the adjacent blocks, the label may transit to a next part with a low probability and remain invariable with a high probability; and when w=T/5, the label may transit to a next part with a high probability.

Function f measures the compatibility between two labels. Different types of dependencies have different significance on the result. f k(x, y) denotes the compatibility between two labels in the k th type of dependency. λ_k denotes the weights of the k th type of dependency.

Parameter estimation

Once the function f k and g k are given, the parameter estimation aims to determine the parameter $\theta = (\lambda_1, \lambda_2, ...; \mu_1, \mu_2, ...)$ from the training dataset D={(x(i), y(i)), i=1, 2, …N}. The conditional model is trained discriminatively based on the Conditional Maximum Likelihood (CML) criterion, which maximizes the log conditional likelihood.

$$\theta^* = \arg\max_{\theta} \sum_{i=1}^{N} \log P(y^i \mid x^i; \theta). \tag{3}$$

Gradient ascent algorithm is applied to calculate the θ^*. Calling ts the unnormalized Ps(Y|X), θ s is the parameters in component Ps. To calculate the gradient $d\theta_s$ of θ s, the following learning rule is used:

$$d\theta_s = \left\langle \frac{\partial \log t_s}{\partial \theta_s} \right\rangle_{P_0(Y|X)} - \left\langle \frac{\partial \log t_s}{\partial \theta_s} \right\rangle_{P_\theta(Y|X)} \tag{4}$$

Where P0(Y|X) is the experiential distribution defined by D, and P θ (Y|X) is the model distribution. However, calculating expectations under the model distribution is difficult because of the normalization factor Z. According to the structure between nodes, belief propagation[23] is applied to calculate the gradient.

Inference

Inference aims to find the optimal class label configuration y given an image and gets a probability P(y|x). BP is an inference method proposed by Pearl[24] to efficiently estimate Bayesian beliefs in the network by the way of iteratively passing messages between the adjacent nodes. In our framework, the graph structure used to Infer is a network with loops. Therefore, loopy belief propagation (LBP)[24] is used to obtain an approximate inference. m_{ij} is supposed to be the message sent to node j from node i. There are two variations of the BP algorithm: sum-product and max-product. The former obtains the approximate MPM inference, while the latter does the MAP one. The max-product is selected and m_{ij} is written as:

$$m_{ij}(y_j) \leftarrow \max_{x_i} g_i(y_i) f_{ij}(y_i, y_j) \prod_{k \in \eta_i \backslash j} m_{ki}(y_i) \tag{5}$$

When all the confidence degrees are converged, the confidence of node i is defined as:

$$b(y_i) \propto g_i(y_i) \prod_{j \in \eta_i} m_{ji}(y_i) \tag{6}$$

Finally, the inference result is obtained by searching the label which maximizes the confidence b (yi). Thus, an optimal part label and the corresponding probability are obtained.

CRF for face detection

A given image can be applied to the trained CRF model to calculate the probability corresponding to an optimal label. However, it’s not our ultimate goal. To determine whether an image is a face or not, a threshold needs to be determined. When the probability is higher than the threshold, the given image is a face; otherwise, it’s not. The results of the face detection are sensitive to the threshold. The threshold that gives optimal results is called the optimal threshold, which is obtained by training.

The detector only detects a face at a given location and a given scale. To detect the face at any position within an image, the detectors are performed for all possible positions at a given scale. To detect the object at any size, the input image is iteratively resized and the detector is performed to every resized image. The ratio between two consecutive scale levels is a constant selected empirically and equals 1.2. There exist many probabilities which are higher than the threshold over neighboring locations and several scales. Collisions over scales are removed by selecting one with the highest probability.

EXPERIMENTS AND RESULTS

Experimental data and evaluation criteria

The frontal face was collected from the classical face database such as XM2VTS[25], BioID[26] and FERET[27]. 9000 face images were obtained by scaling, rotation and translation, in which 4000 images were for training and 5000 images for testing. In addition, the background image was collected from the internet and 100000 non-face images were obtained. All the images were cropped and resized to 112 x 92 pixel size. By training with face images and non-face images, an optimal threshold differentiating the face and non-face was obtained. In order to reduce the false detection rate, some face-like false positives were selected designedly such as that illustrated in Figure 2.

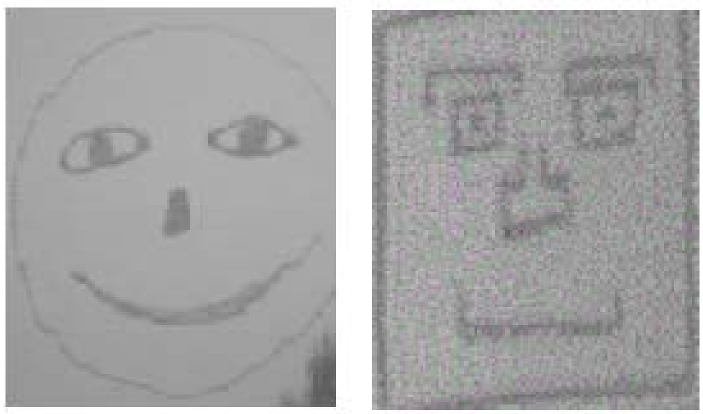


Figure 2: Face-like false positive

The detection method was evaluated in terms of the detection rate and false detection rate. Detection rate denotes the rate of the correctly detected face to all the faces. The false detection rate denotes the rate of the incorrectly detected non-face to all the faces.

Impact of Face segmentation method on Face Detection

An investigation has been made on the influence of the following factors on the detection performance: height of block (L), amount of overlap (p) and number of State (N). Experimental results are illustrated in TABLE 1.

TABLE 1: Impact of face division on face detection

L	P	N	Det.rate (%)	False positive (%)
8	6	5	93.1	15
		6	87.0	14
12	10	5	92.0	10
		6	89.1	9
16	11	5	85.3	10
		6	82.1	9
20	14	5	82.4	9
		6	80.5	8

The results in TABLE 1 show that when L is smaller, the false detection rate and the detection rate are higher. Therefore, we should strick a balance between the two criterions. By analyzing the results, relatively good results can be obtained when L= 12, P=10, N= 5, which is regarded as the optimal segmentation. The trained weights of five parts are 0.08, 0.15, 0.35, 0.30, and 0.12 from top to bottom.

Experiments on different features and different dependencies

In case of the optimal segmentation, one or two kinds of features (e.g. DCT, LBP, DCT and LBP) were extracted and different dependencies were integrated to compare the detection rate (det. rate) and the false detection rate. Experimental results are illustrated in TABLE 2.

As is shown in TABLE 2, adding to a kind of feature or adding to a kind of dependency can improve the detection rate and reduce the false detection rate evidently. The weights of DCT features and LBP features are 0.7 and 0.3 respectively, while the weights of dependencies with w=T/5 and 1 are 0.9 and 0.1 respectively.

The low-frequency DCT coefficients mainly reflect the global structure of the block while the LBP reflects the texture. The experimental results indicate that the weight of DCT is greater than that of LBP. This demonstrates that the global structure feature plays a more important role in face detection. However, the texture feature may be crucial to face recognition[23].

TABLE 2: Comparison of results on different feature and dependency

feature	dependency			
	W_1=1		W_1=1 and W_2=T/5	
DCT	Det.rate	62.1%	Det.rate	70.6%
	False positibes	10%	False positibes	9%
LBP	Det.rate	48.3%	Det.rate	62.1%
	False positibes	12%	False positibes	12%
DCT+LBP	Det.rate	80.3%	Det.rate	90.1%
	False positibes	9%	False positibes	6%

The weight of dependency with w=T/5 is greater than that with w=1, and the reason may be that distance of w=T/5 means a transition from a part to another; However, the distance of w=1 only signifies that the adjacent regions are similar. These results may have practical significance for other part-based object detection.

Experiments on occlusion

From the FDDB dataset[28] we selected some images containing faces with occlusion such as illustrated in Figure 3, some of which were correctly detected and some were not. It is discovered that the more faces with horizontal occlusion can be detected than those with vertical occlusion. This may be determined by face segmentation method. For the face is divided from top to bottom, transverse occlusion only destroys a part of rectangle blocks and vertical occlusion may destroy almost all rectangle blocks.

Experimental results on CMU/MIT dataset

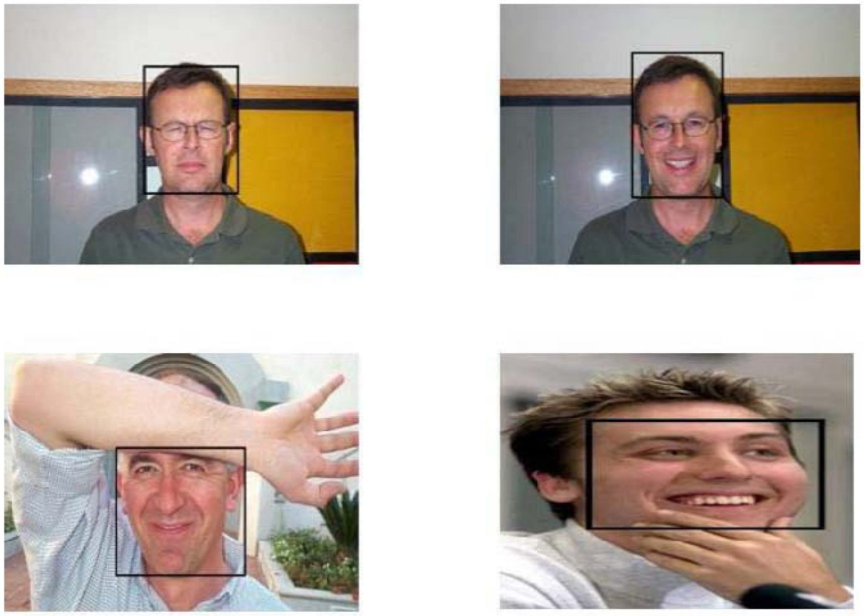
According to Section 4.2 and 4.3, the face detection was performed in terms of separation with L=12, P=10, integrating two kinds of dependencies and extracting two kinds of features: DCT and LBP in each block.

Experiments have been carried out on the CMU/MIT testing set introduced by Rowley[31]. The first version of this testing set contained 23 images with a total of 155 very low resolution faces (it is referred as Dataset 1 in TABLE 3). The complete set contains 130 images with 507 faces (Dataset 3 in TABLE 3). However, some of these annotated faces are manually drawn and they are counted as false

Detections in some publications. To address this ambiguity, some papers only consider 123 images with 483 faces (Dataset 2 in TABLE 3). The three versions of the dataset are tested in this paper. Figure 4 shows some detection results on images of this testing set.

TABLE 3: Detection results on CMU/MIT dataset

Database	#images	# faces	Det.Rate	#false positives
Dataset 1	23	155	88.2%	16
Dataset 2	123	483	94.5%	45
Dataset3	130	507	92.1%	49



(a)

(b)

Figure 3: (a) Some correctly detected faces with occlusion. (b) Some un-detected faces with occlusion

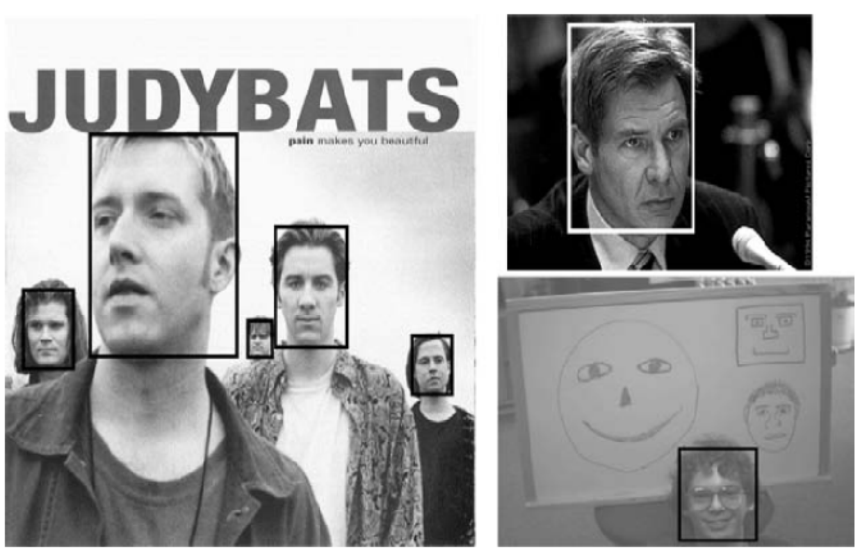


Figure 4: Detection results on some images of CMU/MIT dataset

CMU/MIT database is a classic face detection testing database set. Many researchers have conducted experiments on it. TABLE 4 lists some detection results in the state-of-the-art methods for detecting human faces in the CMU/MIT database. As seen from the results, the method this paper present achieve a higher detection rate and lower false detection rate than Rowley[29], Villa and Jones[17].

TABLE 4: The results of classic face detection method in the CMU/MIT test dataset

Method	D.R.	N.F.
Rowley[29]	90.5	570
Shneiderman etc.[32]	94.4	65
Villa and Jones etc.[33]	91.4	50
Julien etc.[34]	91.7	50
S. Paisitkriangkrai[35]	90.5	50
this paper	94.5	45

CONCLUSIONS AND FUTURE RESEARCH

This paper presents a face detection method based on a CRF framework. In this framework various kinds of features and multiple dependencies are integrated. The weights for each kind of feature, each part and each kind of dependency are achieved by training, for different factor have different effects on face detection. The experimental results illustrate that adding features, adding dependency and optimal segmentation can improve detection rate and the robustness in case of occlusion in face detection. This conclusion may have practical significance for object detection and recognition.

In a future work we propose to improve the detection performance from the following aspects:

1. Adding to the transverse correlation and detecting faces in a 2D CRF framework.
2. Increasing the speed of detection by building multi-level detection and rapidly eliminating non-face region.

ACKNOWLEDGMENT

This paper is partially supported by the Natural Science Foundation of China (Grant Nos. 60902083).

REFERENCES

[1] E.Hjelmås, B.K.Low; Face detection: A survey[J], Computer vision and image understanding, **83(3)**, 236-274 **(2001)**.

[2] M.H.Yang, D.Kriegman, N.Ahuja; Detecting faces in images: A survey[J], Pattern Analysis and Machine Intelligence, IEEE Transactions on, **24(1)**, 34-58 **(2002)**.

[3] C.Zhang, Z.Zhang; A survey of recent advances in face detection[R], Tech.rep., Microsoft Research **(2010)**.

[4] Jianguo Wang, Tieniu Tan; A new face detection method based on shape information, IEEE Trans.Pattern Recognition letters, **21(6-7)**, 463-471 **(2000)**.

[5] J.Meynet, V.Popovici, J.P.Thiran; Face detection with boosted gaussian features, Pattern Recognition, **40(8)**, 2283–2291 **(2007)**.

[6] B.Heisele, T.Serre, T.Poggio; A component-based framework for face detection and identification, International Journal of Computer Vision, **74(2)**, 167-181 **(2007)**.

[7] G.Z.Yang, T.S.Huang; Human detection in a complex background, Pattern Recognition, **27(1)**, 53-63 **(1994)**.

[8] F.S.Samaria, S.Young; HMM-Based Architecture for Face Identification, Image and Vision Computing, **12(8)**, 537-543 **(1994)**.

[9] A.V.Nefian, M.H.Hayes III; Face detection and recognition using hidden Markov models, International Conference on Image Processing, **1**, 141-145 **(1998)**.

[10] V.Nefian, M.H. Hayes III; An Embedded HMM-Based Approach for Face Detection and Recognition, Proc. IEEE Int’l Conf., Acoustics, Speech, and Signal Processing, **6**, 3553-3556 **(1999)**.

[11] Boris Epshtein, Shimon Ullman; Feature Hierarchies for Object Classification, Tenth IEEE International Conference on Computer Vision (ICCV'05), **1**, 220-227 **(2005)**.

[12] J.Lafferty, A.Macullum, F.Pereira; Conditional random fields: Probabilistic models for segmenting and labeling sequential data, Eighteenth international conference on Machine Learning (ICML-2001), 282-289 **(2001)**.

[13] Andrew McCallum, Wei Li; Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons, In Seventh Conference on Natural Language Learning (CoNLL) **(2003)**.

[14] Andrew McCallum, Ben Wellner; Conditional models of identity uncertainty with application to noun coreference, In Lawrence K.Saul, Yair Weiss, and L´eon Bottou, editors, Advances in Neural Information Processing Systems 17, Cambridge, 905-912 **(2005)**.

[15] Yan Liu, Jaime Carbonell, Peter Weigele, Vanathi Gopalakrishnan; Segmentation conditional random fields (SCRFs): A new approach for protein fold recognition, In ACM International conference on Research in Computational Molecular Biology (RECOMB05) **(2005)**.

[16] Zhihui Du, Zhaoming Yin, Wenjie Liu, D.Bader; A case study on Conditional Random Fields training algorithm for biological sequence alignment, IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW), pp. 543-548 **(2010)**.

[17] X.He, R.Zemel, M.Carreira-perp inan; Multiscale conditional random fields for image labeling, IEEE Conference on Computer Vision and Pattern Recognition, 695-702 **(2004)**.

[18] M.Y.Yang, W.Forstner; A hierarchical conditional random field model for labeling and classifying images of man-made scenes, IEEE International Conference on Computer Vision Workshops (ICCV Workshops), 196-203 **(2011)**.

[19] Shravya Shetty, Harish Srinivasan; Sargur Srihari, Handwritten Word Recognition using Conditional Random Fields, ICDAR, 1098-1102 **(2007)**.

[20] A.Quattoni, S.Wang, L.P.Morency; Hidden Conditional Random Fields, IEEE Trans.on Pattern Analysis and Machine Intelligence, **29(10)**, 1848-1852 **(2007)**.

[21] H.Derin, H.Elliott; Modeling and Segmentation of Noisy and Tex2 tured Images Using Gibbs Random Fields, IEEE Transactions on Pattern Analysis and Machine Intelligence, **9(1)**, 39-55 **(1987)**.

[22] Ahonen, A.Hadid, M.Pietik¨ainen; Face recognition with local binary patterns. European Conference on Computer Vision, Prague, 469–481 **(2004)**.

[23] J.Pearl; Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann Publishers **(1998)**.

[24] Y.Weiss; Belief propagation and revision in networks with loops, Technical Report MIT A.I.Memo 1616 **(1998)**.

[25] XM2VTS, http://poseidon. csd. auth. gr/M2VTS/index. html.

[26] BioID, http://www. bioid. com/downloads/facedb/index. php.

[27] FERET, http://www.frvt.org/.

[28] Vidit Jain, Erik Learned-Miller; FDDB: A Benchmark for Face Detection in, Unconstrained Settings. University of Massachusetts, Amherst **(2010)**.

[29] H.A.Rowley, S.Baluja, T.Kanade; Neural network-based face detection. IEEE Trans, Pattern Anal.Mach.Intell., **20(1)** , 23-38 **(1998)**.

[30] Chih-Chung Chang, Chih-Jen Lin; LIBSVM : a library for support vector machines, Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm **(2001)**.

[31] Timo Ahonen, Abdenour Hadid, Matti Pietikainen; Face Description with Local Binary Patterns: Application to Face Recognition, IEEE Trans. on Pattern Analysis and Machine Intelligence, **28(12)**, 2037-2041 **(2006)**.

[32] H.Zhou, X.Li, D.Huang et al.; Detecting hedges scope based on phrase structures and dependency structures[C], Proceedings of 7th International Conference on Natural Language Processing and Knowledge Engineering (NLP-KE), 415-420 **(2011)**.

[33] W.Li, V.Mahadevan, N.Vasconcelos; Anomaly Detection and Localization in Crowded Scenes[J], IEEE Transactions on Pattern Analysis and Machine Intelligence, **36(1)**, 18-32 **(2014)**.

[34] Q.Huang, L.Y.Wu, X.S.Zhang; Cneta: Network alignment by combining biological and topological features[C], Proceedings of IEEE 6th International Conference on Systems Biology (ISB), 220-225 **(2012)**.

[35] A.Shivram, B.Zhu, S.Setlur et al.; Segmentation Based Online Word Recognition: A Conditional Random Field Driven Beam Search Strategy[C], Proceedings of 12th International Conference on Document Analysis and Recognition (ICDAR), 852-856 **(2013)**.