# Mixed Facial Emotion Recognition using Active Appearance Model and Hidden Conditional Random Fields

*by* Dewi Yanti Liliana

International Journal of Pure and Applied Mathematics
Volume 118 No. 18 2018, 3159-3167
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
Special Issue

ijpam.eu

# Mixed Facial Emotion Recognition using Active Appearance Model and Hidden Conditional Random Fields

[1]Dewi Yanti Liliana, [2]T. Basaruddin, [3]M. Rahmat Widyanto
[1] PhD Student of Faculty of Computer Science, Universitas Indonesia
Email:dewiyanti.liliana@tik.pnj.ac.id
[2]Faculty of Computer Science, Universitas Indonesia
Email: chan@cs.ui.ac.id
[3] Faculty of Computer Science, Universitas Indonesia
Email:widyanto@cs.ui.ac.id

## Abstract

Automatic emotion recognition through facial expression analysis is an emerging topic on affective computing and social signal processing which gains more attention. The development on this field has manyinfluences in human life related to the machineinteraction understand human to meet human needs.Existing research on emotion recognition focuses on recognizing basic emotions (happy, sad, surprise, disgust, fear, and angry), but less effort has been done for mixed emotion recognition due to its complexity. The challenge on mixed emotion recognition as a combination of basic emotions are still widely open and has not much been explored. We proposed a combination of Active Appearance Model (AAM) as a facial feature extraction framework and Hidden Conditional Random Field (HCRF) as a temporal classifier with hidden states as a model of mixed facial emotion recognition. The experiment on an arranged temporal CK+ dataset and our own mixed emotion dataset shows an improvement in accuracy rate compare to our previous methods, original CRF and SVM-CRF classifier.

**Keywords:** mixed emotion recognition, facial expression analysis, active appearance model, hidden conditional random fields.

## 1.Introduction

One of the biggest research challenge in intelligent machine is on exploring human behavior and how they interact with their environment. Social Signal Processing (SSP) is a research area which aims at bridging human and computer interaction through the analysis of nonverbal signal (social signal) such as facial expression, intonation, gesture, and body position [1].Research on machine learning to analyze and synthesize humanemotion become a leading issue. Recently, facial expression analysis has become a research focus because face gives moreinformation about human emotion. Facial expression conveys 55% informationof human emotion in a daily communication compared to another social signal[2].

Automatic facial expression recognition is amongst the hot topic in SSP and still facing many challenges. This field has many potential applications, such as in human computer interaction, human emotion analysis, robotics, biometric recognition, lie detection, special needs treatment, entertainment, and education. Particularly in human emotion analysis, human has variety of emotions from facial expressions including basic and mixed emotions. Mixed emotion from facial expression is somehow uneasy to be recognized even by another human. Designing a machine learning algorithm to recognize human complex emotions is by far still an open challenge[3].

This research encompasses two aspects; understanding mixed emotion in terms of human facialemotion and modeling the recognition of mixed emotion from facial expression images using machine learning classifier.We collaborate Active Appearance Model (AAM) as a framework which produces facial landmark and the Hidden Conditional Random Fields (HCRF) as a sequence classifier with hidden layer which yields mixed emotion's class label for sequence of images input. we organize our research paper into six sections. We start with introduction about the problem, related works on human facial emotion recognition, the design of the proposed work, experiment and results, and conclusion.

## 2.Related Works

Facial expression analysis and recognition aims to classify the emotion on a face image. Referring to Sumathi et al., facial expression analysis involves three stages: facial acquisition, feature extraction, and facial expression recognition [3].Facial acquisition separates facial area from non-facial area in an image.Feature extraction extracts facial area into feature vectors and facial expression recognition classifies feature vectors into appropriate emotion classes.Face acquisition is the first step to detect face either from a single image or on a set of images. Many methods applied for face detection including edge detection, boosting techniques, and automatic segmentation [4], [5].Wu (2015) used Restricted Boltzmann Machine to detect face points [6].

Two common feature extraction technique for facial expression recognition are geometric and appearance feature extraction. Geometric features used facial points location (e.g. eye corner, lip corner, etc.) or face component's shape (e.g. eye, eyebrow, mouth, etc.). Appearance features used texture of face which is robust to the illumination variation [7]. The frequently used methods are Gabor wavelet and Local Binary Pattern (LBP) because it tolerates illumination changing andhas simple computation [8]. Both features have strength and weakness. Appearance features might be good at handling the variation of illumination but it is less sensitive to a small shift. This weakness can be covered by geometric features which use the face area to describe changes in it. On the other side, geometric feature is lack of texture information. Good feature representation can be obtained by combining the strength of various features, this is called hybrid features [9], [10]. Active Appearance Model is a framework which combines appearance and texture features to produce facial landmarks [11].Feature extraction plays an important role in facial expression recognition because at this stage the feature vector is generated as an input to the next stage and determines the recognition result.

Facial expression recognition is the last stage in facial expression analysis. There are two categories, frame-based and sequence-based. Frame-based expression recognition does not use temporal information in an input image. The input can be either a static image or separated image frames. Mostly, SVM is used as learning methods for facial expression recognition because it gives a high accuracy rate [12], [13]. Other methods are also applied with good results, such as multilayer perceptron (MLP) [8], convolutional neural network [14], [15], and fuzzy clustering [16]. Sequence-based recognition uses temporal information of the input sequence to recognize the expression of one or more frames.Sequence-based is more challenging because it used sequence classifier such as HMM and CRF, and it needs different modelling and representation [17]–[19].The existing facial expressions recognition using temporal information used sequential classifiers such as Hidden Markov Model [17], [18] and Conditional Random Fields [19]–[21], and HCRF [19], [22].

Majority research on emotion recognition are classifyingsix classes of basic emotion (happy, sad, surprise, fear, disgust, and angry). Somehow, human real emotions are more complex and widely involving mixed emotion. Research on mixed emotion recognition is still on initial phase. Du et. al (2104) developed compound emotion recognition, a combination of basic emotions in one image[23].Their research identified 21 classes of mixed emotions which consisted of 6 classes of basic emotions and 15 classes of compound emotions from still images and applied SVM for classification. Our previous works has been successfully classified 12 classes of mixed emotion on temporal image frames SVM-CRF classifier [24]. Our proposed work is using hidden structure of HCRF to repair the sequence classification process and to increase the accuracy rate.

## 3.Mixed Facial Emotion Recognition

We design temporal mixed facial emotion recognition using the combination of AAM and HCRFon a sequence of images. The recognition phase consists of three steps as in Figure 1,facial feature extraction and sequence mixed emotion recognition. The input to the system is a sequence of image frames which contains basic emotion in each frame. This sequence of basic emotions forms a mixed emotion class as result of the recognition. AAM extracts facial landmarks of each image framebased on the mean shape of reference image. Subsequently, HCRF classified the sequence of basic emotion classes into the respected mixed emotion class.

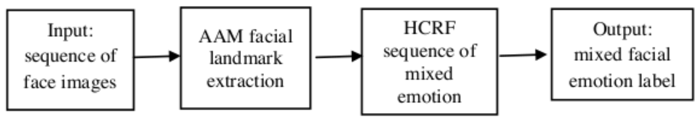


Figure 1. Mixed facial emotion recognition phases

### 3.1. Mixed Emotion

Emotion is an important element in social interactions since it shows a response in communication. Psychologist Ekman and Friessen define basic emotion as a separate discrete emotions which significantly differ from one another[25]. They categorized six classes of basic emotions: happy, sad, angry, surprise, disgust, and fear. They also organized basic emotion into a hierarchical of emotion family based on characteristics similarity such as expression similarity, psychological activity, and the event that trigger those emotions to occur. These characteristics distinguish emotion classes from one another.

Mixed emotion is an affective experience which involves two basic emotions, usually the opposing valence emotions such as happy and sad. The definition implies that there is emotion states transition over the time[26]. Moreover, an emotion can elicit other emotion instantly to develop an emotional experience. This is shown by the combination of facial expression in a short time duration[27]. Basic emotions are fundamental emotions in life which can be used as a guide to explain more complex other emotion. We used 12 classes of mixed emotion, as in Du, et.al. (2014) [23]: happily-surprised, happily-disgusted, sadly-surprised, sadly-fearful, sadly-disgusted, sadly-angry, fearfully-surprised, disgustedly-surprised, angrily-surprised, fearfully-disgusted, fearfully-angry, and angrily-disgusted.

### 3.2. Facial landmarks by AAM

Active Appearance Model is a statistical template matching model which captures facial characteristics and yields facial landmarks or facial points as an output[28]. AAM is a framework that works with the principle of feature extraction and combine both of shape and texture features on a face image. Those features are extracted using PCA to create a face model which consists of several landmarks or coordinate points. Facial points are scattered on specific face region: jaw, eyebrows, eyes, nose, and mouth.

AAM framework consists of several procedures which sequentially adjust the shape to the mean shape by using Procrustes method. AAM is applied on each face image and resulting a pairwise vector of facial point coordinates. AAM result becomes the input of the next phase, the recognition phase. At the next step, SVM is employed as a classifier of basic emotion class. SVM is chosen because of its high accuracy rate on classifying the multiclassification problem.

### 3.2. HCRF for Mixed Facial Emotion Recognition

Conditional random fields (CRFs) is an undirected graphical model which predicts the label of a given sequenceby maximizing the conditional probability of sequence variables[29]. CRF has been known outperformedothersequential classifier such as Hidden Markov Model (HMM), and Maximum Entropy Markov Model (MEMM).The improvement of CRFs was made by Quattroni, et.al. who proposed the hidden-states of unobserved latent variables or so-called Hidden CRF (HCRF) [30]. HCRFused hidden variables to model the latent sub-structure of the problems. HCRFs discriminatively defines a joint distribution of the class label and the latent variable label conditioned by the observations. Just like hidden state in HMM, hidden variables in

CRFs allows the use of training data without explicit label provided. HCRF is used in classification problem with hidden-state learning on local features.

Mixed facial emotion recognition task has an underlying sub-structure which builds the mixed emotion class. This sub-structure consists of sequence of basic emotions with undirected manner. The discriminative approach is suitable for this sequential case, because it discriminates one basic emotion class against other classes. Moreover, the problem can be modeled as a hidden structure using discriminative HCRF to predict the sequence label or mixed emotion label. Previous research on mixed emotion recognition using CRF has shown promising results, but the model isunsuitable because it does not incorporate hidden variables which is naturally occurs by the problem itself [24].

HCRF is a discriminative model with hidden states and it is well-suited for the mixed emotion recognition problem. The original HCRF by Quattoni used hidden state to capture the spatial dependency between hidden object parts[30]. We model the HCRF to the sequence of input where the underlying sub-structure captures the temporal dependency between image frames.We design the HCRF sequence classifier for each observation $x$ of a classlabel $\in Y$, where x is a vector of $m$sequence of observations $x = \{x_1, x_2, \dots x_m\}$. In our problem, $x_i$is an observation of facial image frame $i$. $x_i$ is represented as a feature vector $x_i \in R$. The conditional probability of a mixed emotion class label from a sequence of image facial expression frames is computed using HCRF model:

$$P(y|x,\theta) = \sum_h P(y,h|x,\theta) = \frac{\sum_h e^{\psi(y,h,x;\theta)}}{\sum_{y' \in Y, h \in H^m} e^{\psi(y',h,x;\theta)}} \quad (1)$$

where$h$ = $\{h_1, h_2, \dots h_n\}$is a set of hidden states in the model and each $h_i \in H$ captures the underlying structure of each six classes of basic emotion.Ψ is a potential function using parameter Θ to measure the label y, observation$x$ and hidden states. In regular CRF, there is a certain label and no hidden states $h$, assuming that $h$ is observable class.We used gradient ascent learning to gain the optimal value for $\theta$ by maximizing log likelihood of the data where $\theta^* = argmax_\theta L(\theta)$.

## 4. Experiments and Results

We test our proposed AAM HCRF model and compare it results with CRF model, and SVM-CRF model on a modified CK+ dataset [31] as well asour own made mixed emotion dataset. Each input sequence is limited to seven image frames consist of two different basic emotion in arbitrary order that form the mixed emotion. The output label comes from12 classes of mixed emotion: sadly surprised, sadly fearful, sadly-disgusted, sadly angry, happily-surprised, happily disgusted, disgustedly surprised, angrily-surprised, angrily disgusted, fearfully disgusted, fearfully angry, and fearfully surprised.We arrange sequences of image from CK+ dataset which contains facial images of six classes of basic emotion (happy, sad, surprise, fear, angry, and disgust).We used 306 images as training data and create 80 sequences of testing data. We run experiment on our own dataset consists of 12 mix emotion classes from 15 subjects;Indonesiancitizens with different ethnicity (Javanese, Batak, Sundanese, Malay, and Chinese). Our own mixed emotion dataset gives 270 training data and 48 sequences of testing data. Figure 2 gives example of images sequences from both dataset.

(a)

(b)

Figure 2. Sequence of image frames (a) CK+ dataset (b) mixed emotion dataset

AAM produces vector of 68 landmarks on each input image. Figure 3 shows the AAM result implemented using MATLAB programming. This vector becomes the input to the HCRF sequence classifier. Figure 4 is a screen capture of the HCRF classification result over an instance of image sequence.

Figure 3. AAM facial landmarks

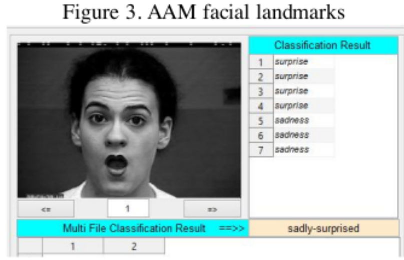


Figure 4.HCRF classification result

Table 1. Experiment results

| Method | Accuracy (%) | |
|---|---|---|
| | CK+ | own |
| CRF | 82.65 | 77.5 |
| SVM-CRF | 90.48 | 83.33 |
| HCRF | 93.19 | 85.71 |

We compare our proposed HCRF classifier with the original CRF and SVM-CRF classifier from our previous works [24]. Table 1 summarizes the result. HCRF gains higher accuracy rate compare to the other methods, both on CK+ dataset and own mixed emotion dataset. On CK+ HCRF accuracy rate is 93,19%, while in own dataset is 85.71%. The proposed HCRF classifier has improved the accuracy rate than the original CRF as well as SVM-CRF methods.

## 5. Conclusion

We extend the common research on facial expression analysis for basic emotion recognition into a more real-complex mixed emotion recognition in a temporal dimension by collaborating Active Appearance Model and Hidden Conditional Random Fields. To the best of our knowledge, there is no reporting use of AAM-HCRF for mixed emotion recognition previously. AAM is a framework which produces facial landmark dynamically and HCRF works as a sequence classifier with hidden states that yields mixed emotion's class label for sequence of images input. The experiment shows that our proposed HCRF model outperforms the existing sequential classifier model and the accuracy rate is 93.19%. Our next will focus on increasing the performance of HCRF by modifying the internal structure as well as optimizing the learning algorithm.

## References


[1] A. Vinciarelli, H. Salamin, and M. Pantic, "Social Signal Processing: Understanding Social Actions through nonverbal behaviour analysis," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshop CVPR Workshop 2009*, vol. 3, no. 231287, pp. 42–49, 2009.

[2] Z. Ming, A. Bugeau, J.-L. Rouas, and T. Shochi, "Facial Action Units intensity estimation by the fusion of features with multi-kernel Support Vector Machine," in *Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on*, 2015, vol. 6, pp. 1–6.

[3] S. C P, "Automatic Facial Expression Analysis A Survey," *Int. J. Comput. Sci. Eng. Surv.*, vol. 3, no. 6, pp. 47–59, Dec. 2012.

[4] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *ArXiv 12070580*, pp. 1–18, 2012.

[5] M. Pantic and I. Patras, "Detecting facial actions and their temporal segments in nearly frontal-view face image sequences," *2005 IEEE Int. Conf. Syst. Man Cybern.*, vol. 4, pp. 3358–3363, 2005.

[6] Y. Wu and Q. Ji, "Discriminative Deep Face Shape Model for Facial Point Detection," *Int. J. Comput. Vis.*, vol. 113, no. 1, pp. 37–53, 2015.

[7] M. Rasoulzadeh, "Facial Expression recognition using Fuzzy Inference System," vol. 1, no. 4, pp. 1–5, 2012.

[8] A. Gudi, H. E. Tasli, T. M. D. Uyl, and A. Maroulis, "Deep Learning based FACS Action Unit Occurrence and Intensity Estimation," vol. 2013, 2015.

[9] F. De la Torre and J. F. Cohn, "Facial expression analysis," *Vis. Anal. Hum.*, pp. 377–410, 2011.

[10] P. A. D. Martins, "Active appearance models for facial expression recognition and monocular head pose estimation," University of Coimbra, 2008.

[11] L. Wang, R. Li, and K. Wang, "A novel automatic facial expression recognition method based on AAM," *J. Comput.*, vol. 9, no. 3, pp. 608–617, 2014.

[12] X. Liu, "Video-based face model fitting using Adaptive Active Appearance Model," *Image Vis. Comput.*, vol. 28, no. 7, pp. 1162–1172, 2010.

[13] M. Kim, "Conditional ordinal random fields for structured ordinal-valued label prediction," pp. 378–401, 2013.

[14] B. Fasel and J. Luettin, "Automatic facial expression analysis: a survey," *Pattern Recognit.*, vol. 36, no. 1, pp. 259–275, Jan. 2003.

[15] D. A. Pitaloka, A. Wulandari, T. Basaruddin, and D. Y. Liliana, "Enhancing CNN with Preprocessing Stage in Automatic Emotion Recognition," *Procedia Comput. Sci.*, vol. 116, pp. 523–529, 2017.

[16] D. Y. Liliana, M. R. Widyanto, and T. Basaruddin, “Human emotion recognition based on active appearance model and semi-supervised fuzzy C-means,” in *2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 2016, pp. 439–445.

[17] J. Zraqou, W. Alkhadour, and A. Al-Nu’aimi, “An efficient approach for recognizing and tracking spontaneous facial expressions,” *2013 Second Int. Conf. E-Learn. E-Technol. Educ. ICEEE*, pp. 304–307, 2013.

[18] M. F. Valstar and M. Pantic, “Fully Automatic Recognition of the Temporal Phases of Facial Actions,” *IEEE Trans. Syst. Man Cybern. Part B Cybern.*, vol. 42, no. 1, pp. 28–43, Feb. 2012.

[19] B. Hasani and M. H. Mahoor, “Spatio-Temporal Facial Expression Recognition Using Convolutional Neural Networks and Conditional Random Fields,” *ArXiv Prepr. ArXiv170306995*, 2017.

[20] M. H. Siddiqi, R. Ali, A. M. Khan, Y. Park, and S. Lee, “Human Facial Expression Recognition Using Stepwise Linear Discriminant Analysis and Hidden Conditional Random Fields,” vol. 24, no. 4, pp. 1386–1398, 2015.

[21] L. V. D. Maaten and E. Hendriks, “Action unit classification using active appearance models and conditional random fields,” vol. 13, 2012.

[22] K.-Y. Chang, T.-L. Liu, and S.-H. Lai, “Learning partially-observed hidden conditional random fields for facial expression recognition,” in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 533–540.

[23] S. Du, Y. Tao, and A. M. Martinez, “Compound facial expressions of emotion,” *Proc. Natl. Acad. Sci.*, vol. 111, no. 15, pp. E1454–E1462, Apr. 2014.

[24] D. Y. Liliana, C. Basaruddin, and M. R. Widyanto, “Mix Emotion Recognition from Facial Expression using SVM-CRF Sequence Classifier,” 2017, pp. 27–31.

[25] P. Ekman, “An argument for basic emotions,” *Cogn. Emot.*, vol. 6, no. 3–4, pp. 169–200, May 1992.

[26] R. Berrios, P. Totterdell, and S. Kellett, “Eliciting mixed emotions: a meta-analysis comparing models, types, and measures,” *Front. Psychol.*, vol. 6, p. 428, 2015.

[27] K. Oatley and P. N. Johnson-Laird, “Cognitive approaches to emotions.,” *Trends Cogn. Sci.*, vol. 18, no. 3, pp. 134–140, Mar. 2014.

[28] Xinbo Gao, Ya Su, Xuelong Li, and Dacheng Tao, “A Review of Active Appearance Models,” *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 40, no. 2, pp. 145–158, 2010.

[29] J. Lafferty and A. Mccallum, “Conditional Random Fields : Probabilistic Models for Segmenting and Labeling Sequence Data Conditional Random Fields : Probabilistic Models for Segmenting and,” vol. 2001, no. June, pp. 282–289, 2001.

[30] A. Quattoni and S. Wang, “Hidden conditional random fields,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1848–1853, 2007.

[31] Y.-L. Tian, T. Kanade, and J. F. Cohn, “Recognizing action units for facial expression analysis,” *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. CVPR 2000 Cat NoPR00662*, vol. 1, no. 2, pp. 1–19, 2001.

## Authors Biography

**DewiYantiLiliana** is a Ph.D student at Faculty of Computer Science, Universitas Indonesia. She received her Master Degree on Computer Science from Universitas Indonesia, and bachelor of Computer Science from SepuluhNopember Institute of Technology, Indonesia. Her research topic is focusing on emotion recognition from facial expression analysis through computer vision and machine learning modeling.

**T. Basaruddin**is a Professor of Computer Science at Universitas Indonesia. He received his PhD and Master degree from Manchester University, UK, and bachelor degree on mathematics from Gajah Mada University, Indonesia. His research interests are on Social Signal Processing, Machine learning, Numerical Analysis, and Computational Intelligence.

**M. RahmatWidyanto**is a Doctor of Computer Science at Universitas Indonesia. He received his PhD and Master degree from Tokyo Institute of Technology, Japan, and bachelor degree on Computer Science from Universitas Indonesia. His research interests are on Artificial Intelligence, Soft Computing, and Fuzzy System.

# Mixed Facial Emotion Recognition using Active Appearance Model and Hidden Conditional Random Fields

ORIGINALITY REPORT

| 13% | 3% | 8% | 8% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | Submitted to Liverpool John Moores University<br>Student Paper | 3% |
| 2 | Submitted to Coventry University<br>Student Paper | 1% |
| 3 | Submitted to Universitas Diponegoro<br>Student Paper | 1% |
| 4 | Desy Komalasari, M. Rahmat Widyanto, T. Basaruddin, Dewi Yanti Liliana. "Shape analysis using generalized procrustes analysis on Active Appearance Model for facial expression recognition", 2017 International Conference on Electrical Engineering and Computer Science (ICECOS), 2017<br>Publication | 1% |
| 5 | Hidayanto, Achmad Nizar, Muhammad Azani Hasibuan, Putu Wuri Handayani, and Yudho Giri Sucahyo. "Framework for Measuring ERP Implementation Readiness in Small and Medium Enterprise (SME): A Case Study in | 1% |

Software Developer Company", Journal of Computers, 2013.
Publication

| | | |
|---|---|---|
| 6 | Ariadna Quattoni. "", IEEE Transactions on Pattern Analysis and Machine Intelligence, 10/2007<br>Publication | 1 % |
| 7 | studfile.net<br>Internet Source | 1 % |
| 8 | Submitted to University of Warwick<br>Student Paper | 1 % |
| 9 | N E Mamonto, H Maulana, D Y Liliana, T Basaruddin. "Multimedia Content Development as a Facial Expression Datasets for Recognition of Human Emotions", IOP Conference Series: Materials Science and Engineering, 2018<br>Publication | 1 % |
| 10 | www.hindawi.com<br>Internet Source | 1 % |
| 11 | Submitted to Middlesex University<br>Student Paper | 1 % |
| 12 | www.coursehero.com<br>Internet Source | 1 % |

| | | | |
|---|---|---|---|
| Exclude quotes | On | Exclude matches | < 1% |
| Exclude bibliography | On | | |