OBJECT LOCALIZATION AND ACTIVITIES IDENTIFICATION USING ATTRIBUTE DETAILS IN SMART MEETING ROOMS

Dian Andriana[1,2], Ary Setijadi Prihatmanto[2], Egi Muhammad Idris Hidayat[2], and Carmadi Machbub[2]

[1]Research Center for Informatics, Indonesian Institute of Sciences, Bandung, Indonesia
[2]School of Electrical Engineering and Informatics, Institut Teknologi Bandung, Bandung, Indonesia

ABSTRACT

This paper is concerned with the development of interactive systems for smart meeting rooms. Automated recognition of video events is an important research area. We present an LTL (Linear Temporal Logic) model of basic objects and activities recognition in smart meeting rooms using object attribute details. There are still problems of misrecognizing objects in existing visual recognition methods because lack of enough feature attributive information details. This paper investigates morphological approach to increase recognition accuracy using variability in a limited area of moving object using object attribute details. The proposed methods are also compared to popular and recent methods of visual object and event recognition.

KEYWORDS

Events Recognition and Tracking, Morphological Feature Characteristics

1. INTRODUCTION

This paper is concerned with design of automated interactive systems for smart meeting rooms. Design of automated visual recognition of video events is an important research area to support interactive systems in a very early stage of development. Events and behaviours can be modelled in the context of related visual recognized objects. Previous works [1][2][3] are still unable to visually localize speaker of the meeting rooms and need the help of audio features to identify the speaker. Other researchers [4][5][6] explained smart meeting rooms modelling but did not focused on visual events and behaviours which are important features of smart room systems. This paper aims to focus on visual events or behaviours in the smart meeting room systems including visual localization of the speaker. Research done by [1] uses predefined chair and speaker position. However, in the real situation without predefined scenarios, problems still occur in correctness of object identification due to various object position and illumination [7]. In addition to the face features, the speaker or the participants identification can also be done from other personal attributes such as clothes texture, ties, hair, headscarves, etc. Events and behaviours are also identified visually as the speaker entering the room or approaching the podium, talking, leaving the podium, etc. The participants' interactivity through dialogue can also be identified visually.

Natarajan Meghanathan et al. (Eds) : AIAA, WiMNeT, DPPR - 2018
pp. 63–76, 2018. © CS & IT-CSCP 2018 DOI : 10.5121/csit.2018.81606

The object detection using the popular HOG-SVM (Histogram of Oriented Gradients – Support Vector Machines) method still has problems of correctness due to various object position and illumination [7]. Human visual recognition algorithm has been developed and shows success only in controlled environment featuring face, iris, human action and behaviour [8][9]. Face detection algorithm has been developed and shows success in featuring facial landmarks such as corners of the eyes, the tip of the nose, the mouth, the eyebrows, and the face boundaries using regressing trees [10]. The work describes specific face landmarks in detail features or characteristics such as the eyes, eyebrows, nose, and mouth. However, the facial landmarks method is still not able to describe enough information of facial features compared to human perception capabilities of face recognition. There are still problems with human identification using face and body features while they are moving in pose variation [7]. Many statistical and machine learning methods need huge amount or training data taken from human perception or knowledge about objects [10], but it is still incomplete collected training data [11]. Besides statistical methods, syntactical methods include hierarchical, relational, structural, and morphological methods have been developed for face recognition [12][13][14][15]. Structural hierarchical and relational methods have been done for high level abstraction of image objects modelling, but still have problems for low level implementation [12][13]. Many researchers have developed various human activities and behaviour recognition, such as walking, sitting, bending, and some sport activities, and less work of person identification that does the activities [15].

Personal identification is not only using face, but also can use other personal attributes such as hair, hat, headscarves, body posture, or clothing. Personal attributes can be detected using HOG-SVM, for example the textile textures in clothing. Khan [16] combines colour attributes and shapes and learned using HOG-SVM. Works by Nurhaida [17] use SIFT for static textile images, and less movement of the person wearing the textile. Work by Reddy [18] uses LBP and GLCM combined with KNN and SVM to extract features. Kalantidis [19] uses SIFT and LBP for clothing recognition in different appearances.

In visual object detection, object features has also been investigated through SIFT, SURF, BRISK, BRIEF, FREAK, AKAZE, and ORB [20][21][22][23][24] which are based on corner detector as the best features. They work for simple objects with simple background but fails in more complex objects and complex background. The corner-based methods result in basically random match features which are far from correctness in matching object pairs, and need extended areas from the corner points, for example, using curves for describing shapes. The corner points only do not describe shapes and position of object features.

The popular method HOG-SVM has high accuracy in describing face landmarks using thousands of mean-shapes from provided samples classified by regression tree using fit scores. This provides difference or tolerance values from the mean shapes [10] [25], and able to work with simple shapes / curves, but still have problem with sharp curves / non-simple curves. LBP method [8] can be applied in different size of pixels because of histogram size normalization. However, LBP and HOG-SVM cannot be applied in small size of pixels describing small features. They require more or less similar size of pixels and make distribution of grey level without maintaining the pixel positions. GLCM [18] requires too specific and overfitting pattern with lack of generalization or tolerance due to object position and pose changes. HOG SVM in [10] [25] with high accuracy still has ambiguity results which can be eliminated using additional pixels information processing using pixel grey level curve distance evaluation such as Manhattan distance. Problem with illumination are solved using HOG-SVM, but more precise result revalidated using grey level pixel values. Solving ambiguity in HOG-SVM method basically give primal guidance that all objects can be detected using provided samples by their appearances.

Related objects can elaborate more complex problem in scene understanding in wider context of objects, for example in recognition of events, activities or behaviours of moving human objects,

such as human activities or behaviours. In this paper we use lecture and conference activities in a smart meeting room as a study case of events recognition. Images processed from videos which are sampled from courtesy of Youtube [26][27].

2. MODELLING BASIC OBJECTS AND ACTIVITIES OF SMART MEETING ROOM

Personal visual recognition performance using face and body postures can decrease because of movement in different poses [7]. To better recognized and discriminate personal face and body characteristics, we can use attributes of the human object such as the eyes, nose, mouth, hair, headscarf, hat, ties, fabric cloth texture, etc. [28][29][30][31][32][33]. After personal recognition, related events should be recognized in videos by tracking consecutive events or activities. For events or activities recognition, formal models are used for reasoning. In paper by [34] video events are modelled using Petri nets and formalized using LTL (Linear Temporal Logic) formulas to provide guidance for programming implementation.

Basic activities that can be recognized visually in lecture / conference activities in the meeting room include the speaker identification using attributes details, different attributes of the participants, the speaker and the participant activities, layout of display, podium, tables and chairs related to speaker and participants position. This paper focuses on visual recognition without neither predefined position of the camera nor the speaker. Room layouts such as round tables arrangements as recognized objects are included in the model, for example it is useful for controlling dynamic movement of cameras or microphones in smart rooms. The speaker recognition is the central of the system, and recognized not only from the face, but also from detail attributes and behaviours. In this paper we focus on detail attributes, behaviours of the speaker and the participants of the conference, and the room layouts. For example, we use semi round tables layout as depicted in Figure 1, and we model the semi round tables layout in LTL formulae as follows.

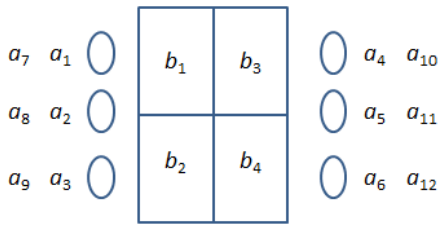

Figure 1. An example of semi round tables arrangement

$$\mathbf{GF}s \tag{1}$$

$$s \vDash \mathbf{F}a_1 \wedge \mathbf{F}a_2 \wedge \mathbf{F}a_3 \ldots \wedge \mathbf{F}a_{12} \tag{2}$$

where

s = satisfied condition of semi round tables arrangement
a1 = the first position of detected person / chair facing right
a2 = the second position detected person / chair facing right
a3 = the third position detected person / chair facing right
a4 = the fourth position of detected person / chair facing left
a5 = the fifth position detected person / chair facing left

a6 = the sixth position detected person / chair facing left
a7 = the first position of detected person / chair facing table
a8 = the second position of detected person / chair facing table
a9 = the third position of detected person / chair facing table
a10 = the fourth position of detected person / chair facing table
a11 = the fifth position detected person / chair facing table
a12 = the sixth position detected person / chair facing table
b1= b2 = b3 = b4 = four small tables
or replaced by b5= one big table.

$$a_4 \vDash (\mathbf{F}a_{41} \wedge \mathbf{F}a_{42} \wedge \mathbf{F}a_{43}) \vee (\mathbf{F}a_{44} \wedge \mathbf{F}a_{45} \wedge \mathbf{F}a_{46}) \quad (3)$$

where

a_{41} = x-coordinate left position of the person / chair a_1< x-coordinate left position of the table b_1
a_{42} = y-coordinate top position of the person / chair a_1≥ y-coordinate top position of the table b_1
a_{43} = y-coordinate bottom position of the person / chair a_1≤ y-coordinate bottom position of the table b_2
a_{44} = x-coordinate left position of the person / chair a_1< x-coordinate left position of the table b_5
a_{45} = y-coordinate top position of the person / chair a_1≥ y-coordinate top position of the table b_5
a_{46} = y-coordinate bottom position of the person / chair a_1≤ y-coordinate bottom position of the table b_5

We define an example of a speaker through talking activity and tie attribute as follows.

$$\Omega = \text{count(talk)} \geq \epsilon \quad (4)$$

$$\neg\, \Omega = \text{count(talk)} < \epsilon \quad (5)$$

$$\Psi = (\mathbf{G}t \wedge (\Omega \mathbf{R}\, d)) \quad (6)$$

where t = the person is wearing a tie and d = counting mouth movement (talking) as described in Figure 2. $\Omega$is a condition when the counting d exceeds a number ϵ which is an assumption of a number of mouth detected opening which is also based on position differences between detected upper and lower lips from theprevious frame-to-frame movement, and also based on counting of the second order differences of the position differences between the detected upper and lower lips as a representation of talking mouth movement. In equation (7) and Figure 3, we also define an example of an interactive participant which does not have a speaker attributes, probably raises hand and stands up for question talks.

$$\mathbf{G}\neg\psi \wedge \text{Fh} \wedge \mathbf{X}((\mathbf{F}c) \vee (\Omega \mathbf{R}\, d)) \quad (7)$$

where

$\neg\psi$ = that person is not a speaker (maybe a participant)
c= the person is standing up (probably eventually)
d = counting mouth movement (talking)
Ω = count(talk) ≥ϵ_1
h= the person raise hand

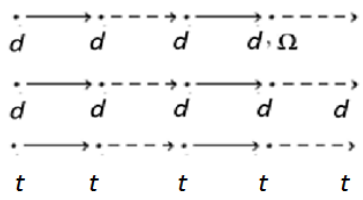

Figure 2. LTL diagram of talking counting and tie attribute

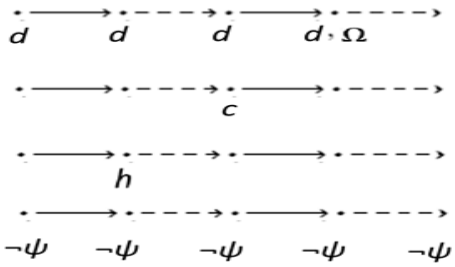

Figure 3. LTL diagram of participant interactivity

In Figure 4, we define an example of a speaker detected with tie attribute and talking activities. We also define other general personal detail attributes such as hair, hat, headscarf, and clothing textile patterns.

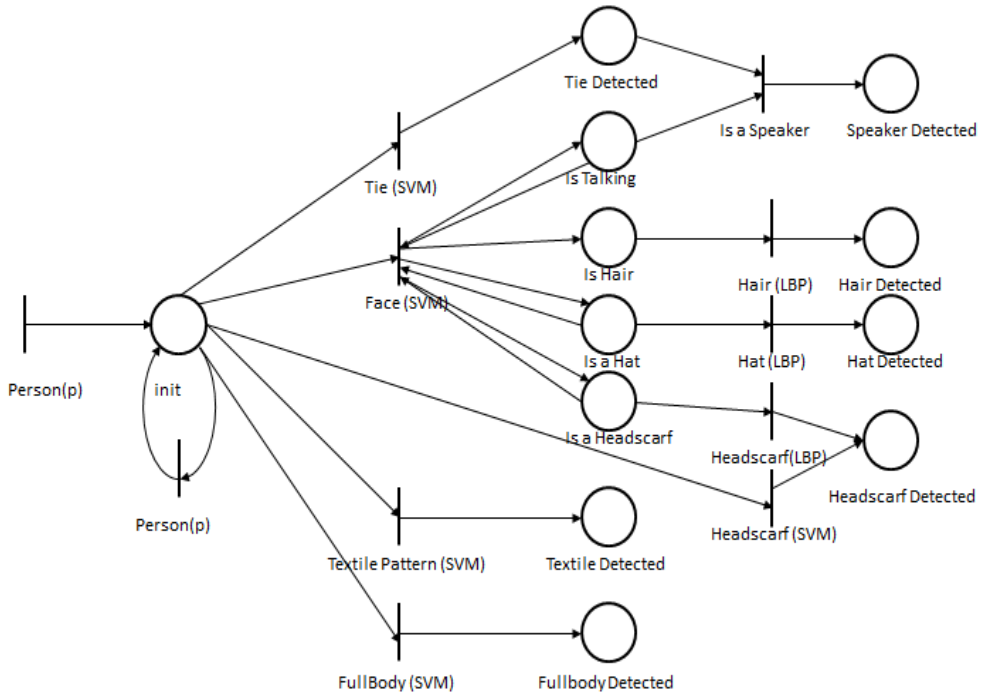

Figure 4. Petri nets diagram of personal detail detection

In Figure 5, we define a scenario of a person movement towards a detected podium, being near podium, talks, and then move away after finish talking. Figure 6 and Figure 7 describe in further details about a person movement detected from position coordinate changes and being near podium coordinates, and eventually detected as a speaker.

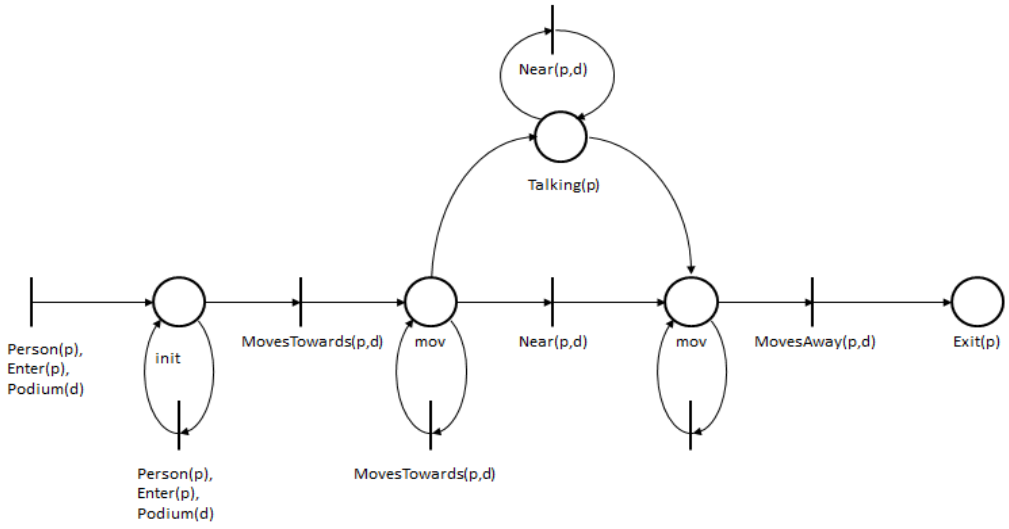

Figure 5. Petri nets diagram of speaker scenario

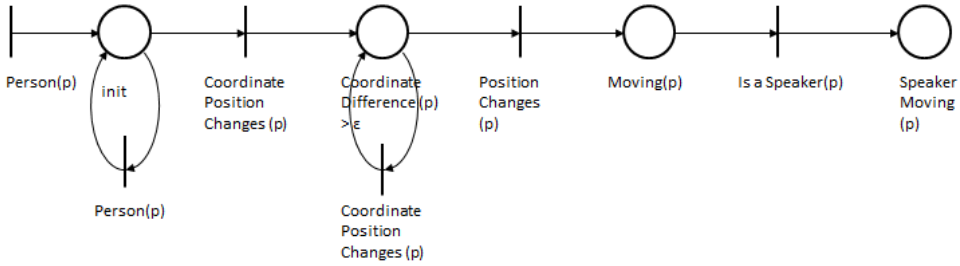

Figure 6. Petri nets diagram of moving speaker

Figure 8 also describes a speaker detected from collected score of checking detected tie, being near podium coordinates, and talking activities from detected lips moving coordinates.

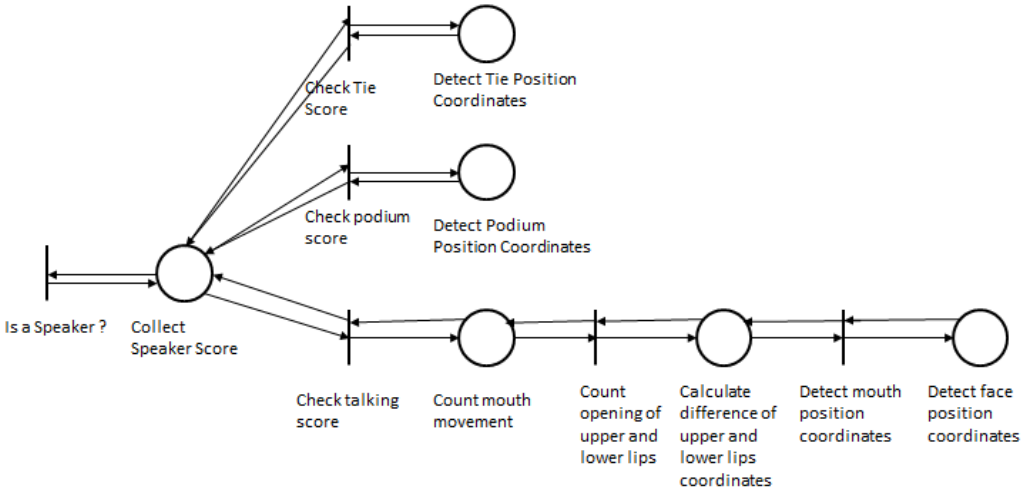

Figure 7. Petri nets diagram of speaker near podium

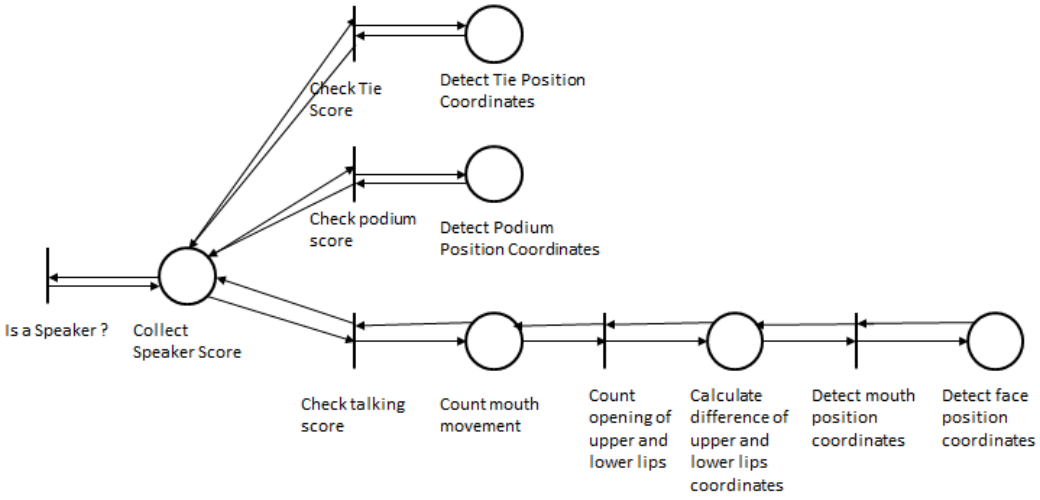

Figure 8. Petri nets diagram of defining speaker properties

3. DETAIL ATTRIBUTE DETECTION USING GRADIENTS

We use an alternative of human recognition to recognize hair texture on moving person in the smart meeting room. In case of small size of the object region of interest (ROI), for example for far objects from camera which implies smaller size of ROIs, we cannot use HOG-SVM due to size of small size area. We cannot also use LBP due to value similarities which are not able to distinguish true and false features. We use Multiple Linear Regression formulas in [31] to describe feature of objects. Examples for identification of the speaker's white hair texture distinguished from the audience's brown hair with guidance of face landmark position are shown in Figure 9 and Figure 10. With position guidance of face landmarks and HOG-SVM object learning [25][35] detection area, we pick hair texture samples from one frame in video and compare them to the other frames. We formulate the pixel grey level value differences below, and then calculate a mean value between Q and R.

$$C_3 < |Q| < C_4\,,\ Q = \{df_t | df_t = f_t - f_{t+1} \cap df_t < 0\} \tag{8}$$

$$C_5 < |R| < C_6\,,\ R = \{df_t | df_t = f_t - f_{t+1} \cap df_t > 0\} \tag{9}$$

The step in Q and R sets are getting the optimum number of pixel points in certain ranges which have negative (or positive) gradients $df_t$with their successor points in the curve of pixel gray levels.

In Table I we use the formulas to describe similarity between sample and tested object features in moving person. We use three ROIs to enable variability of feature areas in moving object, but still in guidance of face landmarks. For example, experiment 1 ROI 2 shows a minimum mean value of pixel gradients Q and R values, and all other experiments in Table I also shows a minimum value at least on an ROI, which is an expected result of the mean formula. In the other side, Manhattan distance calculation fails to show consistent behavior while giving lower value in false features in Table II experiment 4 ROI 1 and higher value in true features in Table I experiment 5 ROI 2. In correct location area, the distance values are expected to be consistently lower in true features and higher in false features. The true and false features are then validated by human eyes, and the mean formula is expected to show consistent expected results.

Table 1. Distance measurement comparison examples for true features.

Experiment	ROI	Mean	Manhattan
1	1	38	2454
	2	5	1301
	3	7	4082
2	1	38	2454
	2	7	1340
	3	3	4092
3	1	13	2688
	2	11	4506
	3	6	4520
4	1	5	2932
	2	10	4412
	3	2	4614
5	1	81	41662
	2	6	36304
	3	41	17976
6	1	73	41640
	2	17	33918
	3	72	21699

Table 2. Distance measurement comparison examples for false features.

Experiment	ROI	Mean	Manhattan
1	1	105	45000
	2	100	46247
	3	85	45045
2	1	105	45027
	2	98	46287
	3	92	45432
3	1	106	45423
	2	104	46321
	3	101	44936
4	1	78	15647
	2	23	29486
	3	40	33221
5	1	82	16664
	2	22	29313
	3	33	32859
6	1	242	66257
	2	118	121455
	3	138	158312

Figure 9. Speaker white hair texture feature, picture is courtesy of Youtube [26]

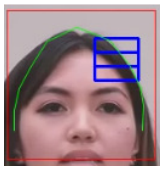

Figure 10. Audience hair texture feature, picture is courtesy of Youtube [26]

4. RESULTS AND DISCUSSIONS

In Table 1 and Table 2, we use mean measurement formula (8) and (9) to clearly distinct true features from false features. Table 1 show mean values on each ROI for each experiment, and for each experiment there exist minimum mean values which are less than 20 for true features and more than 20 for false feature. The minimum mean values occur on at least one of the ROIs of each experiment of true features. In the other side, Manhattan distance measurement cannot distinguish clearly between true and false features.

Choices of segmentation methods and ROI areas are important for defining detail features of object detection. ROI area variability can be extended with guidance of other relatively more reliable features such as face landmarks. There are also possibilities of combining criteria with other attributes such as a person wearing ties. Another challenge is noisy background of the objects. The detection methods can be combined with other features such as colour, but colour attribute is not always useful in outdoor sunny daylight environment which may affect brighter object colour.

Basic objects and activities in smart meeting room are also as the results of objects and activities detection using the LTL formulas and petri nets implementation. Figure 11 shows speaker detection through tie attributes, mouth movement counting, and white hair property in ROIs. In this figure, the mean formula shows white hair property similarity as the lowest value 3 in the bottom ROI position, when Manhattan formula shows inconsistently higher value 4449 compared to 2630 and 1381, also when LBP formula shows indistinctive all zero values in all the ROIs. Figure 12 shows an audience with no tie attributes, and no white hair property since the lowest mean value is 10 which is higher than 3 as the value of white hair property, when Manhattan formula also shows inconsistently higher value compared to other ROIs, and LBP formula shows errors. Figure 13 shows moving person detected, and speaker detected from clothing pattern in the middle ROIs which has the lowest mean value, but LBP formula shows inconsistent higher value in the middle ROI. Finally, in Figure 14 LTL formulas can determine a speaker movement and positioned near podium which is also previously detected by HOG-SVM object detection. Figure 14 also shows the lowest mean value in the bottom ROI, when LBP shows inconsistent higher value in the bottom ROI. Figure 15 is demonstrated by Yolo [36] object detection with correct position of a person, chairs, and a table, but it still misrecognizes a working desk as a dining table. Figure 15 also shows indistinctive similar LBP values for all three ROIs, when the mean value shows the lowest value in the bottom ROI for clothing pattern similarity. The mean formula can distinct object properties in some cases better than HOG and LBP because it maintains the pixel positions rather than distribution of pixel values, also the mean formula works better than HOG and LBP on small size of picture patches as the object properties.

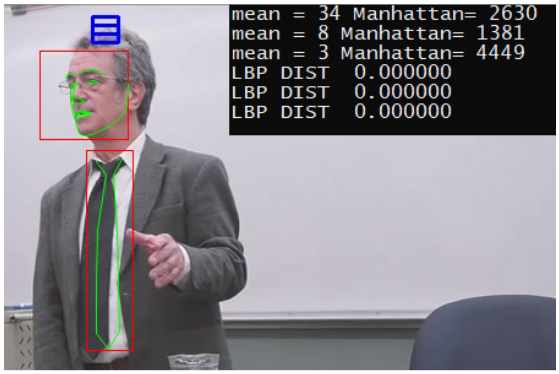

Figure 11. Speaker detected from tie attributes, mouth movement counting, and white hair property in one of the three ROIs, picture is courtesy of Youtube [26]

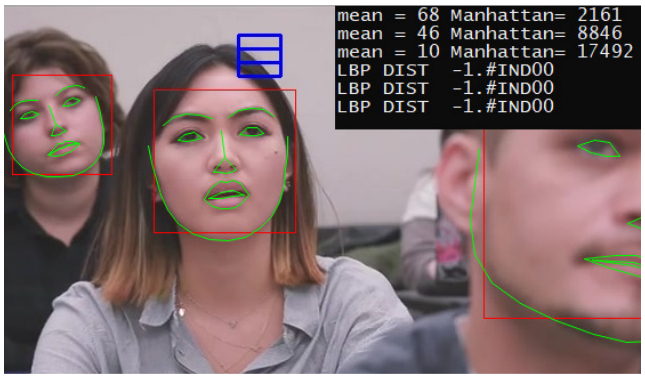

Figure 12. Participant detected from no tie attribute, mouth movement counting, and no white hair property in one of the three ROIs, picture is courtesy of Youtube [26]

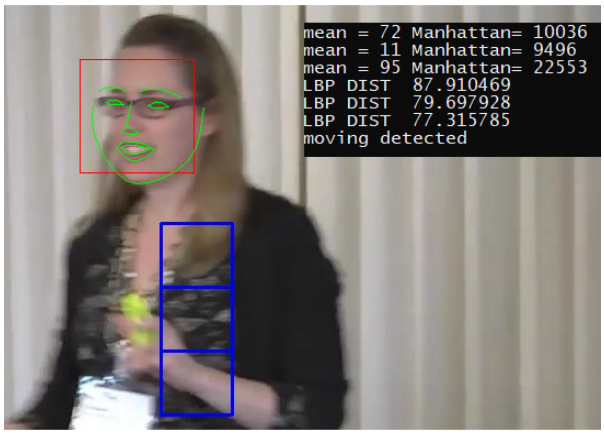

Figure 13. Moving person detected, and speaker detected from clothing pattern in one of the three ROIs, picture is courtesy of Youtube [26]

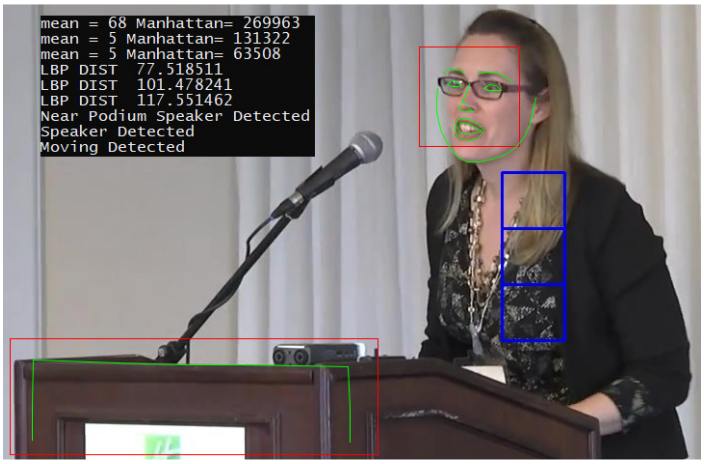

Figure 14. Moving speaker detected near podium, and speaker detected from clothing pattern in one of the three ROIs, picture is courtesy of Youtube [27]

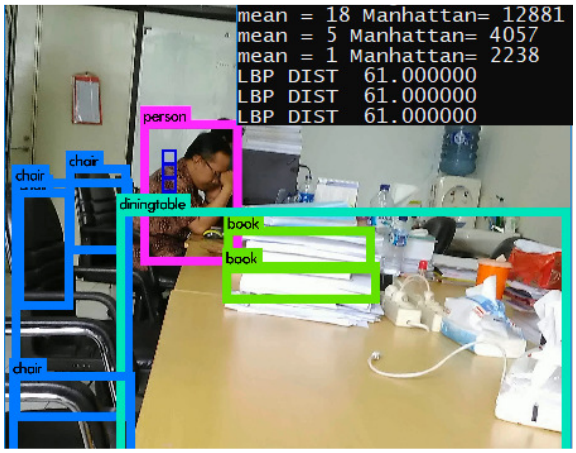

Figure 15. Person, table and chair detected positions using Yolo [36], and a person detected from clothing pattern in one of the three ROIs

5. CONCLUSIONS

This paper has shown a LTL model and its implementation on visual objects and activities recognition in smart meeting rooms. We have also shown that optimum number of pixel points in certain ranges which have negative (or positive) gradients of grey level pixel values can be used to distinct features to increase attribute detail recognition. There are still challenges to distinct more features to increase accuracy of moving object and event recognition. Another possibility is using variability of the gradient sequence.

ACKNOWLEDGEMENTS

We would like to thank Institute of Technology Bandung, Indonesian Institute of Sciences, Indonesian Ministry of Research, Technology and Higher Education especially Saintek Scholarship program to support this research.

REFERENCES

[1] Andrey Ronzhin, Alexander Ronzhin,& Viktor Budkov, (2011) "Audiovisual Speaker Localization in Medium Smart Meeting Room", 2011 8th International Conference on Information, Communications & Signal Processing, Singapore, pp1-5.

[2] Zhao Li, Thorsten Herfet, Martin Grochulla,& Thorsten Thormählen, (2012) "Multiple Active Speaker Localization Based on Audio-Visual Fusion in Two Stages", IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), Hamburg, pp1-7.

[3] Ivan Galov, Rustam Kadirov, Andrew Vasilev & Dmitry Korzun, (2013) "Event recording in Smart Room", 13th Conference of Open Innovations Association (FRUCT), Petrozavodsk, pp20-28.

[4] L. D. Tran et al, (2016) "A Smart Meeting Room Scheduling and Management System with Utilization Control And Ad-Hoc Support Based On Real-Time Occupancy Detection", IEEE Sixth International Conference on Communications and Electronics (ICCE), Ha Long, pp186-191.

[5] Simon Mayer, Nadine Inhelder, Ruben Verborgh, Rik Van de Walle, &Friedemann Mattern, (2014) "Configuration of smart environments made simple: Combining visual modeling with semantic metadata and reasoning"; International Conference on the Internet of Things (IOT), Cambridge, MA, pp61-66.

[6] Maik Wurdel, (2011) "An Integrated Formal Task Specification Method for Smart Environments"; Logos Verlag.

[7] Vishal M. Patel, Jaishanker K. Pillai, & Rama Chellappa, (2011) "Image and Video-Based Biometrics"; T.B. Moeslund et al. (eds.), Visual Analysis of Humans, Springer-Verlag London, pp437-454.

[8] Timo Ahonen, Abdenour Hadid, &Matti Pietikainen, (2004) "Face Recognition with Local Binary Patterns," Computer Vision – ECCV, pp469–481.

[9] Javier Galbally, Sebastien Marcel, &Julian Fierrez, (2014) "Biometric Antispoofing Methods: A Survey in Face Recognition", IEEE Access,Vol. 2, pp1530-1552.

[10] Vahid Kazemi& Josephine Sullivan, (2014) "One Millisecond Face Alignment with an Ensemble of Regression Trees", CVPR '14 Proc. of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp1867-1874.

[11] Stefanos Zafeiriou, Cha Zhang, &Zhengyou Zhang, (2015) "A Survey on Face Detection in The Wild: Past, Present And Future",Comput. Vis. Image Underst., vol. 138, pp1–24.

[12] Laura Antanas, Paolo Frasconi, Fabrizio Costa, Tinne Tuytelaars, & Luc De Raedt, (2012) "A Relational Kernel-based Framework for Hierarchical Image Understanding", SSPR'12 / SPR'12 Proc. of the 2012 Joint IAPR Int. Conf. on Structural, Syntactic, and Statistical Pattern Recognition, pp171-180.

[13] Laura Antanas, Martijn van Otterlo, José Oramas M., Tinne Tuytelaars & Luc De Raedt, (2012) "A Relational Distance-based Framework for Hierarchical Image Understanding", Int. Conf. on Pattern Recognition Applications and Methods, Vol. 2, pp206–218.

[14] Robert Olszewski, (2001) "Generalized Feature Extraction for Structural Pattern Recognition in Time-Series Data", Pittsburgh PA: Carnegie Mellon University.

[15] Hua-Tsung Chen, Yu-Zhen He, Chun-Chieh Hsu, Chien-Li Chou, Suh-Yin Lee, &Bao. Shuh P. Lin, (2014) "Yoga Posture Recognition for Self-training", Int. Conf. on Multimedia Modeling (MMM 2014), pp496-505.

[16] Fahad Shahbaz Khan, Rao Muhammad Anwer, Joost van de Weijer, Andrew D. Bagdanov, Maria Vanrell, & Antonio M. Lopez, (2012) "Color Attributes for Object Detection", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp3306-3313.

[17] Ida Nurhaida, Ary Noviyanto, Ruli Manurung, &Aniati M. Arymurthy, (2015) "Automatic Indonesian's Batik Pattern Recognition Using SIFT Approach",Procedia Computer Science, Vol. 59, pp567-576.

[18] R. Obula Konda Reddy, B. Eswara Reddy, & E. Keshava Reddy, (2014) "An Effective Glcm And Binary Pattern Schemes Based Classification for Rotation Invariant Fabric Textures",International Journal of Computer Engineering Science (IJCES), Vol. 4 Issue 1.

[19] Yannis Kalantidis, Lyndon Kennedy, & Li-Jia Li, (2013) "Getting the Look: Clothing Recognition and Segmentation for Automatic Product Suggestions in Everyday Photos", In Proceedings of the 3rd ACM conference on International conference on multimedia retrieval (ICMR '13). ACM, New York, NY, USA, pp105-112.

[20] Herbert Bay, Tinne Tuytelaars, & Luc Van Gool, (2006) "Surf: Speeded Up Robust Features", Computer Vision–ECCV, Springer Berlin Heidelberg, pp404-417.

[21] Stefan Leutenegger, Margarita Chli, & Roland Y. Siegwart, (2011) "BRISK: Binary Robust Invariant Scalable Keypoints", Computer Vision (ICCV), IEEE International Conference on. IEEE.

[22] Alexandre Alahi, Raphael Ortiz, &Pierre Vandergheynst, (2012) "Freak: Fast Retina Keypoint", Computer Vision and Pattern Recognition (CVPR), IEEE Conference on. IEEE.

[23] Michael Calonder, Vincent Lepetit, Cristoph Strecha, & Pascal Fua, (2010) "Brief: Binary Robust Independent Elementary Features", Computer Vision–ECCV, Springer Berlin Heidelberg, pp778-792.

[24] Ethan Rublee, Vincent Rabaud, Kurt Konolige, & Gary Bradski, (2011) "ORB: an efficient alternative to SIFT or SURF", Computer Vision (ICCV), IEEE International Conference on. IEEE.

[25] Pedro F. Felzenszwalb, Ross B. Girshick, David McAllester, & Deva Ramanan, (2010) "Object Detection with Discriminatively Trained Part Based Models",IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 32, No. 9.

[26] https://www.youtube.com/watch?v=bGP9JRhQXak

[27] https://www.youtube.com/watch?v=Yo845WG3KH8&t=173s

[28] Dian Andriana, (2013) "Linear Function and Inverse Function with Weight Ratio for Improving Learning Speed Of Multi-Layer Perceptrons Feed-Forward Neural Networks", Proc. IC3INA, pp255-259.

[29] Dian Andriana, (2015) "Contiguous Uniform Deviation for Artificial Neural Network Pattern Recognition",Advanced Science Lett. Vol. 21 No. 2, pp189-191.

[30] Dian Andriana, Carmadi Machbub, &Ary Setijadi Prihatmanto Prihatmanto, (2015) "Opponent Zigzag Movement Model Capture and Prediction in Robotic Soccer", International Conference on Interactive Digital Media.

[31] Dian Andriana, Ary Setijadi Prihatmanto, Egi Muhammad Idris Hidayat, &Carmadi Machbub, (2017) "Contiguous Uniform Deviation for Multiple Linear Regression in Pattern Recognition",Journal of Physics: Conference Series, 801 (1), 012046, 2017.

[32] Dian Andriana, Ary Setijadi Prihatmanto, Egi Muhammad Idris Hidayat, &Carmadi Machbub, (2017) "Combination of Face and Posture Features for Tracking of Moving Human Visual Characteristics",International Journal on Electrical Engineering and Informatics, Vol. 9, Iss. 3, pp616-631.

[33] Gabriela E. Martinez, Olivia Mendoza, Juan R. Castro, A. Rodriguez-Diaz, Patricia Melin, & Oscar Castillo, (2015) "Response Integration in Modular Neural Networks Using Choquet Integral with Interval Type 2 Sugeno Measures", Fuzzy Inf. Processing Society (NAFIPS) held jointly with 2015 5th World Conf. on Soft Comput. (WConSC), 2015 Annual Conf. of the North American, Redmond WA, USA: IEEE, pp1-6.

[34] Piotr Szwed & Mateusz Komorkiewicz, (2013) "Object Tracking and Video Event Recognition withFuzzy Semantic Petri Nets", 2013 Federated Conference on Computer Science and Information Systems, Kraków, pp167-174.

[35] Davis E. King, (2009) "Dlib-ml: A Machine Learning Toolkit",Journal of Machine Learning Research 10, pp1755-1758.

[36] Joseph Redmon, Santosh Divvala, Ross Girshick & Ali Farhadi, (2016) "You Only Look Once: Unified, Real-Time Object Detection", 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, pp779-788.

AUTHORS

DianAndriana is now a PhD student in School of Electrical Engineering and Informatics, the Institut Teknologi Bandung (ITB), where she also completed her bachelor' and master's degree. She is also a researcher at the Research Center of Informatics of the Indonesian Institute of Sciences. Her researches interests include decision support and intelligent systems. She has 5 papers published in Scopus indexed journals and conferences.

Ary Setijadi Prihatmanto graduated with B.E. and M.S. in Electrical Engineering at Institut Teknologi Bandung in 1995 and 1998, and received his PhD in Applied Informatics from Johannes Kepler University of Linz, Austria in 2006. He is an associate professor & lecturer of School of Electrical Engineering & Informatics, Institut Teknologi Bandung since 1997. He is also the president of Indonesia Digital Media Forum since 2009. His main interests are Human-Content Interaction, Computer Graphics & Mixed-Reality Application, Machine Learning & Intelligent System, Intelligent Robotics, and Cyber-Physical System.

Egi Hidayat received the B.Eng. degree in Electrical Engineering from the Institut Teknologi Bandung (ITB) in 2003, M.Sc. degree in Control and Information Systems from University of Duisburg-Essen in 2007, and Ph.D. degree in Electrical Engineering from Uppsala University in 2014. He is now with the Control and Computer Systems Research Division at School of Electrical Engineering and Informatics, ITB. His research interests are mainly within the area of system identification and signal processing.

Carmadi Machbub got Bachelor degree in Electrical Engineering from the Institut Teknologi Bandung (ITB) in 1980, Master degree (DEA) in Control Engineering and Industrial Informatics in 1988, and Doctorat degree in Engineering Sciences majoring in Control Engineering and Industrial Informatics from Ecole Centrale de Nantes in 1991. He is now Professor and Head of Control and Computer Systems Research Division, School of Electrical Engineering and Informatics, ITB. His current research interests are in machine perception and control.