

Combination of Face and Posture Features for Tracking of Moving Human Visual Characteristics

Dian Andriana¹, Ary Setijadi Prihatmanto²,
Egi Muhammad Idris Hidayat², and Carmadi Machbub²

¹Indonesian Institute of Sciences, Research Centre for Informatics,

²Institut Teknologi Bandung, School of Electrical Engineering and Informatics, Indonesia

Abstract: Understanding images by recognizing its objects is still a challenging task. Tracking of moving human and recognition have been developed by researchers but not yet shows enough information needed for recognition. Initially, a tracking process of an object starts with detection and recognition of the object in a static pose and position, and then continues in movement in different poses. Available moving human recognition methods still has error in classification and need a huge amount of examples which may still be incomplete. Human face and body posture characteristics such as size of the eyes, nose, mouth, or fat or thin bodies, are important visual features in different poses for personal identification to increase accuracy of human recognition system, and it is still rare in researches. This paper attempts to describe visual features that best known for human, but hard to be recognized by machines. Curve fitting approaches to face and body posture features are also introduced to capture exact patterns of the features. Body postures are also preprocessed with a Kinect depth camera, and also compared to popular and recent methods of visual object recognition. Finally, we demonstrate our method can be useful for visual object classification. Probabilities of personal identification can be increased by using different poses and characteristics of smaller detail features through body postures and face areas. More detail features will richen comparison data samples for higher recognition accuracy.

Keywords: face body posture recognition and tracking, morphological feature characteristics

1. Introduction

Image processing algorithm is used in controlling devices through visual sensors. Moving objects detection and recognition are useful for object tracking, for example in surveillance cameras for human supervision. Human visual recognition algorithm has been developed and shows success only in controlled environment featuring face, iris, human action and behavior [1-3]. 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 [4]. 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 [3]. For body postures, scene background has been eliminated by using a Kinect v.2 [5] depth camera to isolate the front objects which show a human silhouette who stands in front of the camera.

Many statistical and machine learning methods need huge amount or training data taken from human perception or knowledge about objects [4], but it is still incomplete collected training data [6]. Besides statistical methods, syntactical methods include hierarchical, relational, structural, and morphological methods have been developed for face recognition [7-10]. Structural hierarchical and relational methods have been done for high level abstraction of image objects modelling, but still have problems for low level implementation [7-8]. Many

researchers have developed various human activities and behavior recognition, such as walking, sitting, bending, and some sport activities, and less work of person identification that does the activities [11]. Gait identification has been done by [3], but not yet using the whole body combined with face in changing poses for personal identification. Human perception and knowledge of posture recognition usually comes from morphological characteristics of human bodies. Face also has morphological characteristics that combine together for human identification.

Image data understanding is also a pattern recognition problem. Image objects movements can be considered as pixel values pattern transitions over time that can be represented by transformation functions. Research has been done [9] by approximating a pattern structure of data signal in time-series. This work uses a hybrid method which combines statistical methods for discriminating objects, and structural methods for describing morphological (shape-based) features. Usually, shape of objects in morphological feature extractions are described with edges, curves, and regions. Features extraction techniques include chain code, piece-wise linear regression, and curve fitting. Multiple Linear Regression (MLR) in Choquet integral also used in face recognition improvement to enhance learning result from neural networks [12-13]. In [12] the Choquet integral was used as a mechanism to integrate information resulted from neural networks used to learn Sobel edges and morphological gradients, but still needs more testing and training samples. Simple shapes in [4] is also learned using SVM (Support Vector Machine), but still not able to fully learn to follow given samples. In [14] gray-level morphological operators such as dilation and erosion is used, but still do not show further details of shape characteristics and its transformation functions.

Object features has been investigated through SIFT, SURF, BRISK, BRIEF, FREAK, AKAZE, and ORB [15-19] 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 Kinect depth camera has been successfully subtracted the background from the front objects. 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.

Previous work [20-23] has been used successfully transformation function to model the face objects movements. This paper continues the work in combination of the face and body posture movements by utilization of Opencv Dlib face landmarks [24, 25], Kinect [5] and image contouring in Opencv, and it is compared with SURF methods [11].

The human face naturally have properties such as the face width, the eyes width or height, the nose length, the eyebrows thickness, etc. that can naturally derive face further details description. Head attributes such as hair, hat, or headscarves are also useful properties to describe human identification. Body postures also have some properties such as fat or thin bodies which can be inferred from proportional height and width that can be calculated from maximum and minimum position of body posture silhouette pixels.

This paper aims to propose a probabilistic combination of face and body posture characteristics for personal identification based on accuracy uncertainties of human visual detection and recognition. For example, with the limitation of visual recognition accuracy due to movement with changing pose, illumination, or distance from the camera, there are possibilities of ups and downs of the recognition accuracies. Besides that, further details of object features needs to be investigated to increase the probabilities of accuracies.

The rest of the paper is organized as follows: Section 2 describes examples of probabilistic combination of face and full body posture features for human detection and identification, with probabilities of high and low accuracy of recognition. Section 3 describes feature modelling using morphological curve fitting with Choquet Integral that will be used in body posture contouring explained in Section 4. Section 4 describes examples of further natural and smaller detail to increase accuracy probabilities of moving face and body posture recognition through silhouette contouring and curve fitting. Section 5 concludes the whole paper.

2. Combination of Face and Body Posture Features

Generally, visual object detection can utilize existing popular methods such as HOG, SVM, LBP, or SURF [1, 4, 15]. However, these methods can yield low accuracies in uncontrolled scenarios [3]. For instance, HOG-SVM face and headscarf detection fails within close distance (less than ± 3 meters) human object from a laptop camera, with a certain value of ratio of face and body heights taken from face and body detection. Figure 1 shows a case in a longer camera distance when a machine can detect full body of human, but unable to detect face and headscarf as an example of attribute of human object.

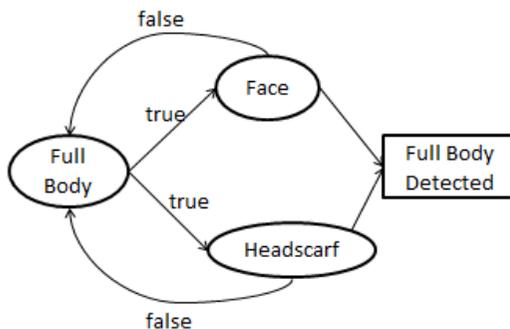


Figure 1. A case of detected body but not detected face and headscarf

In other cases, for example in a closer distance, a machine can detect face and headscarf using HOG-SVM, but shows false positive in detecting full body of human. In this case, to eliminate the false positive results, the full body detection system can be suspended for a while, but still continues with the face detection system. For instance, this is done by checking the ratio of body height compared to the face height. When the ratio is in a certain range of value, the full body detection is suspended, but still continues with face, headscarf, or hat detector. This case is shown in Figure 2.

In Figure 2, the detection processes can be start from any circle symbols in the diagram, and yield results in the rectangle symbols, depends on the transition provided in the circle processes. All processes ran in loop that assumed starts from the full body detection process which includes checking of body and face heights ratio. If a detection process fails, it can return to the full body process and start all over again.

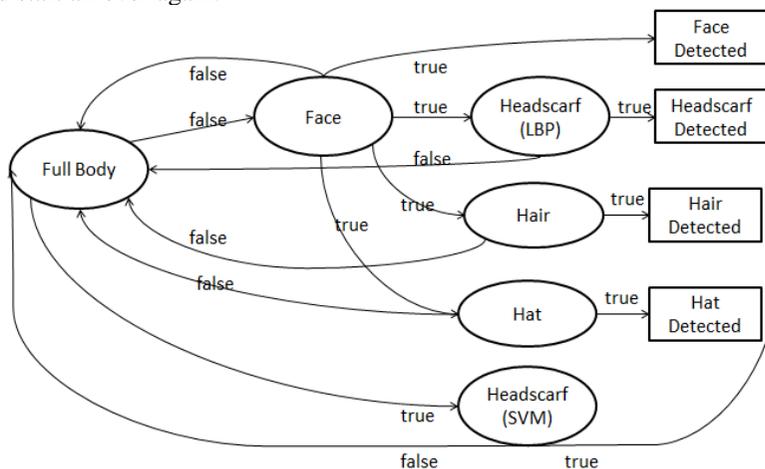


Figure 2. A case of detected face, headscarf, hair, or hat, but not detected full body

Figure 2 shows a case when a full body is not detected in a closer distance, because it is actually half upper body showed in camera. When the face is detected, the face coordinates can guide to head attributes such as headscarf, hair, and hat positions to detect textures using LBP. The positions are necessary to pinpoint the head attributes. The face should be detected first before it can start to detect the headscarf, hair, or hat textures using LBP. But, the headscarf shape detection still can continue using SVM without waiting for the face detection process.



Figure 3. Face and headscarf detection in short distance, and full body detection in long distance from camera

Figure 4 shows examples of women faces with headscarves and hair which also have problems of shapes recognition, because the hair shapes almost similar with the headscarf shapes. While moving in different poses and affected by different illumination, human face cannot accurately be recognized using LBP. Furthermore, LBP work better in smaller areas compared to bigger areas, like showed in Figure 5. Specifically, hair parting is better detected as hair texture to discriminate with headscarf textures. This is because the hair parting textures area has more LBP characterized with mixture of skin and hair textures than the other part of hair textures area. There are also cases where hair shapes are almost similar with headscarf shapes. Figure 5 shows examples of hair parting texture on top of face area detected with texture identified with result in texture ID number 21, and headscarf texture identified with result in texture ID number 42. We also use multiple texture samples which represents different area size of region of interests and different illumination to increase probability of correct recognition. For example, headscarf textures are identified not only with texture ID number 42, but also by texture ID number 42, 62, and 72, and we summarize all those possible textures detected.

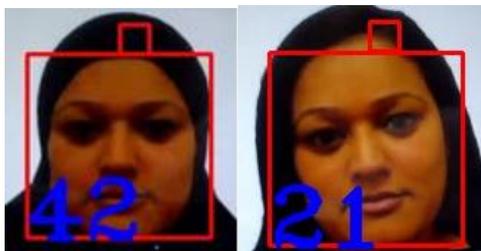


Figure 4. Small LBP areas on top of face area for headscarf and hair detection, face pictures are courtesy of [26]

Further detail features of face should be developed; it can also be combined with body detail features for personal identification. Each person movement in different poses has different visual characteristics. This could be data of characteristics collected for each person and each pose described in Figure 5. For example, face characteristics of a person facing front is different from turning head or body into a certain angle right or left.

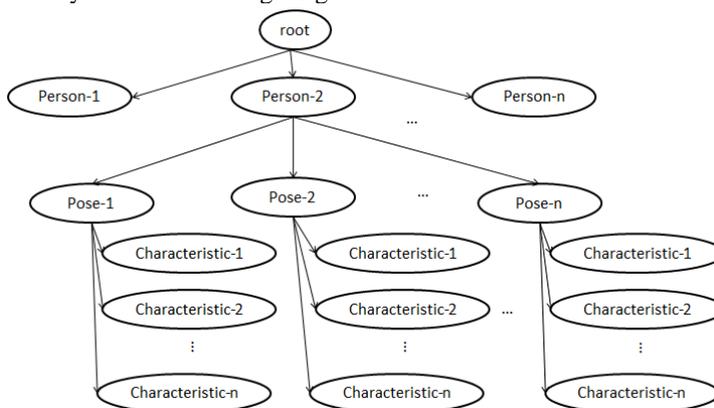


Figure 5. Characteristics of various poses of different persons

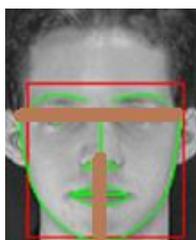


Figure 6. Characteristic of ear-to-ear and nose-chin distance ratio

Examples of face characteristics:

- Ratio of nose length / nose width
- Ratio of nose length / mouth width
- Ratio of nose length / width of eyes area
- Ratio of nose length / face width (ear-to-ear)
- Ratio of nose height (top or bottom) / face width
- Ratio of chin length / face width
- Ratio of eyes height / eyes width
- Ratio of eyebrows height / face width
- Examples of combination of face-body characteristics:
- Ratio of body height / face width
- Ratio of body height / body width
- Ratio of face height / body width
- Ratio of shoulder width / face width

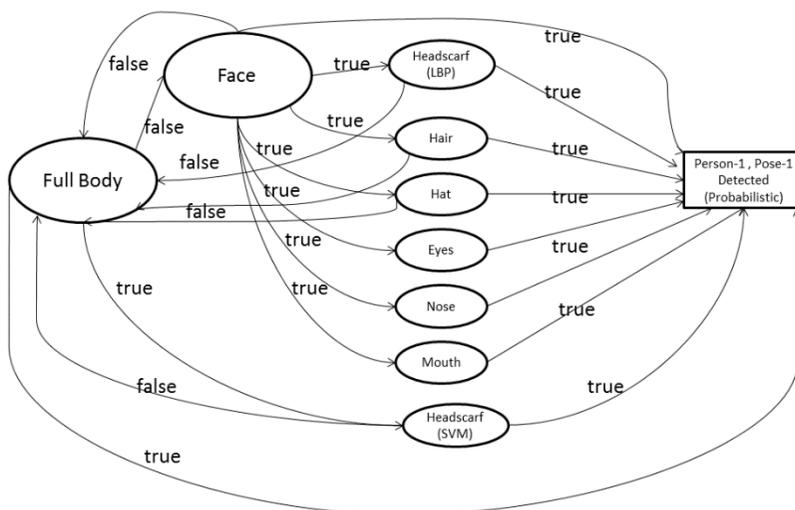


Figure 7. Small LBP in face areas combined with full body recognition to increase probability of personal identification

Recognition performance using the whole face or the whole body areas can decrease because of movement in different poses. In the other hand, detail features from smaller areas are better recognized using LBP than from bigger areas. To better recognized and discriminate personal face and body characteristics, we also use LBP for smaller identified areas such as around the eyes, nose, mouth, hair texture, headscarf, or hat. Figure 7 shows probabilities of personal identification in different poses using characteristics of smaller detail features through body and face areas (with OR switches) to increase probability of personal identification.

3. Feature Modelling Using Choquet Integral

Structural (shape based) detail features description of images can use edge and curves using curve fitting [9]. Choquet Integral was used to approach curve fitting in ref. [27]. Choquet Integral was used to improve face recognition together with Artificial Neural Networks [12, 13]. Choquet Integral formula [28] is described as follows:

$$C_{\mu}(f) = \sum_{i=1}^n w_i \cdot f(c_i) = \sum_{i=1}^n \mu(\{c_i\}) \cdot f(c_i) \tag{1}$$

Choquet Integral is defined as the sum of combination of weights w_i and functions f of criteria c_i , and $\mu(\{c_i\})$ is also defined as the weights of criteria. We can also write the integral as follows, where \bar{y} is an objective value in interval (fuzzy values), \bar{f} is function value in interval, μ is weight, a is scaling parameter vector (a_1, a_2, \dots, a_n) , b is shift parameter vector (b_1, b_2, \dots, b_n) , and c is adjusting constant.

$$\bar{y} = c + \int (a \cdot \bar{f} + b) d\mu \tag{2}$$

Multiple Linear Regression is used to estimate the w_i values used in Choquet integral [11]. For observation variable x_1, x_2, \dots, x_n which is also called predictive attribute, and y as objective attribute, with l sets of data training, described as a matrix below [9]:

x_1	x_2	\dots	x_n	y
\bar{f}_{11}	\bar{f}_{12}	\dots	\bar{f}_{1n}	\bar{y}_1
\bar{f}_{21}	\bar{f}_{22}	\dots	\bar{f}_{2n}	\bar{y}_2
\vdots	\vdots	\ddots	\vdots	\vdots
\bar{f}_{l1}	\bar{f}_{l2}	\dots	\bar{f}_{ln}	\bar{y}_l

(3)

with row component: $\bar{f}_{j1} \ \bar{f}_{j2} \ \dots \ \bar{f}_{jn} \ \bar{y}_j$ and \bar{f}_{ji} is the values from observation variables.

Linear pattern model has also already been developed in Inverse Function Neural Network using linear function as activation function [20] and has been applied in image object pattern recognition for static images [21]. Application in the static image showed capability to model fluctuation of the pixel values. The inverse function Neural Network calculates the weight changes of the neural network nodes using formulae as follows:

$$\sum_{j=0}^n W_{i,j,k+1,t} \cdot x_{i,t} = C_t \sum_{j=0}^n W_{i,j,k,t} \cdot x_{i,t} \quad (4)$$

$$C_t = \frac{g^{-1}(y_{j,t}-p)}{g^{-1}(y_{j,t}-Err_{k,t})} \quad (5)$$

with $w_{i,j,k,t}$ is the connection weight between nodes, and C_t is the ratio of weight changes for the training set t .

We expect the values of curve point differences Δ_k to be relatively almost equal each other along both compared curves [23]. We also define the standard deviation between Δ_k values with mean μ , and the standard deviation are expected to be in a range between certain values C_1 and C_2 . The next 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 [23].

$$\Delta_1 \approx \Delta_2 \approx \Delta_k \approx \dots \approx \Delta_n \quad (6)$$

$$\nabla_t = \begin{cases} |\Delta_t - \Delta_k| < \varepsilon_1 & \text{or} \\ \varepsilon_2 < \Delta_t - \Delta_k < \varepsilon_3 \end{cases} \quad (7)$$

$$C_1 < \sigma = \left(\frac{1}{n} \sum_{t=1}^n (\Delta_t - \mu)^2 \right)^{\frac{1}{2}} < C_2 \quad (8)$$

$$\max_t |P|, P = \{ \Delta_t \mid C_1 < \sigma < C_2 \} \quad (9)$$

$$C_3 < |Q_t| < C_4, Q_t = \{ df_t \mid df_t = f_t - f_{t+1} \cap df_t < 0 \} \quad (10)$$

$$C_5 < |R_t| < C_6, R_t = \{ df_t \mid df_t = f_t - f_{t+1} \cap df_t > 0 \} \quad (11)$$

$$\min_t S, S = \frac{|dQ_t| + |dR_t|}{2} \quad (12)$$

Equation (12) calculates dQ_t and dR_t which are differences between Q_t and R_t in tested points compared to the training points, and we use minimum average value of dQ_t and dR_t which also represents similarities between tested and training points.

4. Methodology and Results

We use face and body posture characteristics model using MLR to learn and test process of recognition. For example, we use MLR to process small areas of face around the right eyes in front facing and turning head 30 degree of three different persons in ORL database of faces from reference [29] as shown in Figure 8. Eyes areas are based on face landmarks detection [4] which has accurately shows positions of the eyes and the eyebrows. Different sizes of areas below the eyebrows first normalized to be similar sizes before MLR processes. Then we take a specific smaller area in the middle of those areas to enable specific discrimination between different persons through specific characteristics of the eyes in gray level pixels. In MLR, gray levels fluctuation is processed without changing the pixels position. The gray level fluctuation can

become eyes characteristics of the persons. Examples of sequence of the eyes gray level characteristics curves are shown in Figure 9, and the results of person recognition through the eyes gray level characteristics are shown in Table I. Table I shows comparison among persons characteristics in calculation of standard deviation values of pixel gray level differences in equation (8) and average values of negative and positive gradients counting differences in equation (12).

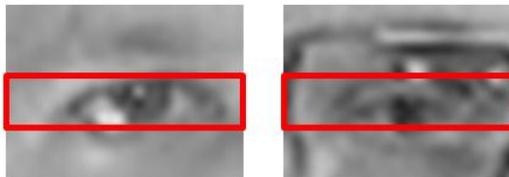


Figure 8. Smaller areas in the middle of eyes areas of two different persons

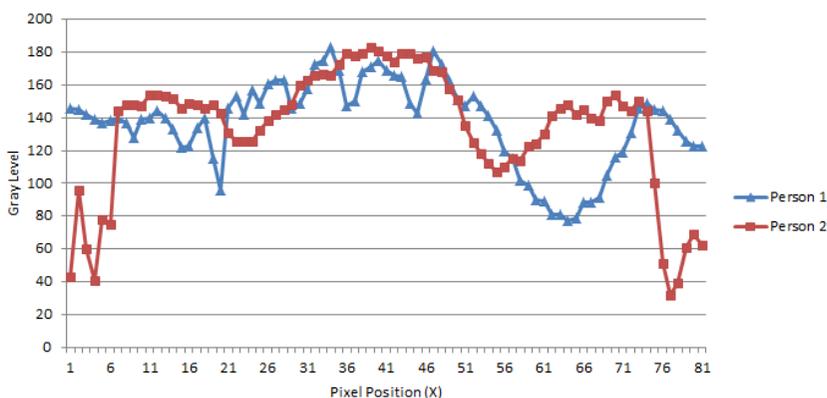


Figure 9. A comparison of gray level sequences in pixel positions

Table 1. Comparison of eyes characteristics gray level curves in Figure 8 and 9

Experiment	Person	Facing	Person	Facing	Eyeglasses	St.Dev	S
1	1	Front	1	Turning	N	17.20	23.5
2	1	Front	1	Front	N	19.41	9.5
3	1	Front	1	Turning	N	38.57	7
4	1	Front	2	Front	Y	37.59	24.5
5	1	Front	2	Turning	Y	19.22	44.5
6	1	Front	4	Front	Y	25.35	69.5
7	1	Front	4	Turning	Y	21.83	14
8	1	Front	4	Front	N	22.03	46.5
9	1	Front	4	Turning	N	19.76	54.5
10	1	Front	4	Front	N	22.05	84.5
11	1	Front	4	Turning	Y	34.79	44.5
12	1	Front	4	Front	N	21.12	42.5
13	1	Front	4	Turning	N	18.54	18

Table 1 shows comparison among three persons characteristics in many different poses, with eyeglasses or not. In experiment 1, person ID 1 is compared to person ID 1 in a different pose, and yields the lowest standard deviation values of pixel gray level difference which shows a correct recognition that the second person has similar characteristics with the first person. Experiment 3 also shows the lowest average values of negative and positive gradients counting difference which also shows correct recognition of the person.

Another example of facial characteristics using LBP processes of smaller areas of face is shown in Figure 10. We use areas around the right eye in front facing and turning head 30 degree

of three different persons, and show results in Table II. The first experiment yields the smallest X^2 result which is a successful recognition of person ID 1 as correct person identification.



Figure 10. Area around the right eye of different persons in different poses

Table 2. Comparison of right eyes LBP characteristics

<i>Experiment</i>	<i>Person</i>	<i>Facing</i>	<i>Person</i>	<i>Facing</i>	<i>Eyeglasses</i>	X^2
1	1	Front	1	Turning	N	2.5
2	1	Front	1	Front	N	5.4
3	1	Front	1	Turning	N	15.3
4	1	Front	2	Front	Y	2.6
5	1	Front	2	Turning	Y	3.1
6	1	Front	4	Front	Y	2.7
7	1	Front	4	Turning	Y	3.5
8	1	Front	4	Front	N	3.8

For body posture, we use silhouettes from background subtraction using Kinect, and then the contour can be extracted using OpenCV showed in Figure 11. We then try to match pairs of silhouettes and also contours using SURF [11] in simulated movement which also include various scales and illuminations. Then we count the correct match compares to the wrong match to calculate the true positive and the false positive rates.

Silhouettes matching using SURF [11] only cannot satisfy accurate object recognition since they results in high false positives. Then the high false positive rate can be lowered using contour of the silhouettes which also result in low true positives.

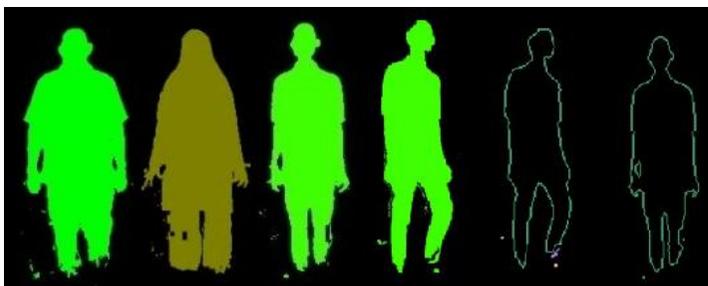


Figure 11. The silhouette and contour examples in matching experiments

The shape of a shoulder can be considered as a curve which is approximated by Multiple Linear Regression in Cartesian coordinates, but first we need to eliminate circular curve by taking only one side of body as shown in Figure 12. We match the shoulder curve pattern pairs of a person facing front and turn left ± 45 degrees and give results with coefficient of determination $R^2 = 0.9330$ in Figure 13, Figure 14 and Table 3.

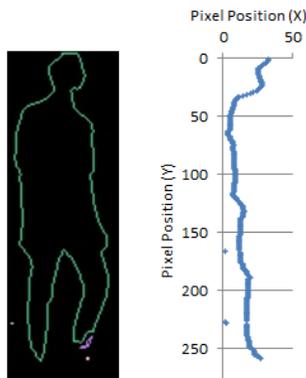


Figure 12. The right outer side of a contour in pixel position coordinates

We also use the contiguous uniform deviation algorithm which is based on values of curve difference in the same range when the test curve is compared to the reference curve [20]. This distance evaluation method is better compared to Mahalanobis distance method [30] because it does not need translation of curve position since it calculates the deviation of differences. Mahalanobis does not work if the curve position is translated to be overlapped each other to enable R^2 calculation. Table 3 shows zeros and almost zeros in all position for Mahalanobis distances. This means the Mahalanobis distances incapable of discriminating between the true and false positions. The contiguous uniform deviation method works in translated as well as non-translated curve positions.

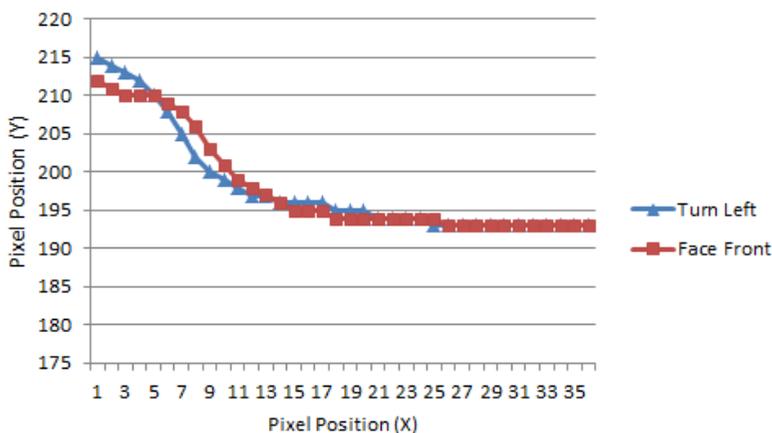


Figure 13. The shoulder shapes curve fitting in for pattern pairs of a person facing front and turning left ± 45 degrees in rotated pixel position coordinates

Table 3. Comparison of different start point of curve

Start Position	R^2	Uniform Deviation	Mahalanobis
24	0.93	1.48	5.01×10^{-14}
25	0.90	1.89	2.55×10^{-14}
26	0.82	2.35	3.47×10^{-14}
27	0.66	2.79	0
28	0.42	3.23	2.76×10^{-14}
29	-0.08	3.65	2.82×10^{-14}
30	-0.94	4.05	0

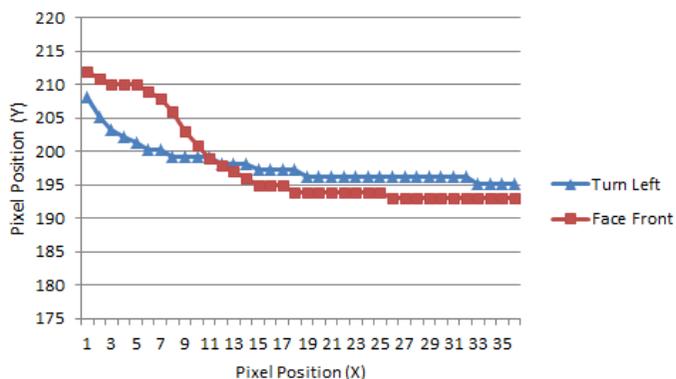


Figure 14. The shoulder shapes in Figure 13 using different start points of curves

The contour curves of body silhouettes also show uniqueness in R^2 and the contiguous uniform deviation in the whole image pixel area. This is useful for object segmentation, for example to isolate the shoulder segments for posture characteristics from the whole image area. In Figure 15 until Figure 18, each figure shows that there is only one single peak of the maximum R^2 along the axis of pixel position, and there is only one single valley of the minimum Standard Deviation (the lowest value) in both start point positions, and we can still see the uniqueness also in all other start positions in the image area.

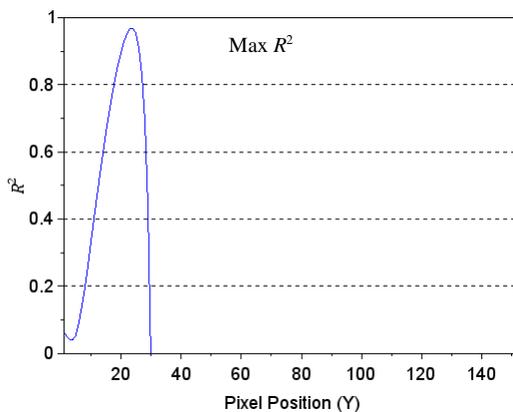


Figure 15. Uniqueness of maximum R^2 for start position 24



Figure 16. Uniqueness of minimum Standard Deviation for start position 24

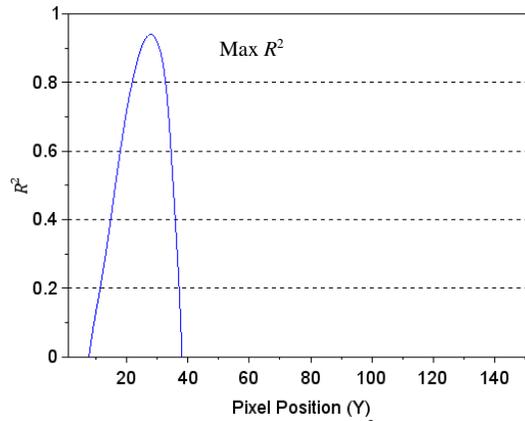


Figure 17. Uniqueness of maximum R^2 for start position 29

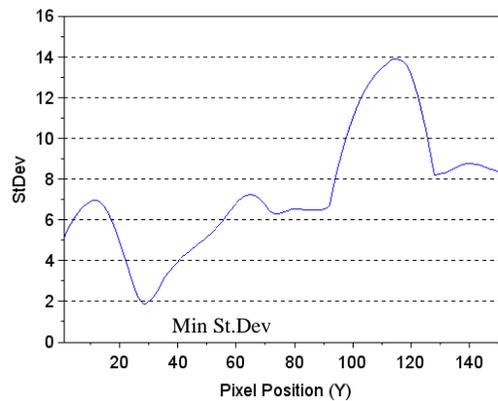


Figure 18. Uniqueness of minimum Standard Deviation for start position 29

We use common characteristics of body postures such as thin and fat, using rectangle of contour borders (ratio of width and height) in Figure 19 which shows the minimum and maximum pixel positions of body contours extracted from the silhouettes. We want to evaluate for both cases: front facing pose and turning ± 45 degrees. For the case of front facing pose, we compare the thin and fat bodies, and see the curve differences using evaluation methods.

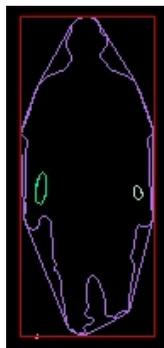


Figure 19. Rectangle of maximum and minimum pixels position of contours

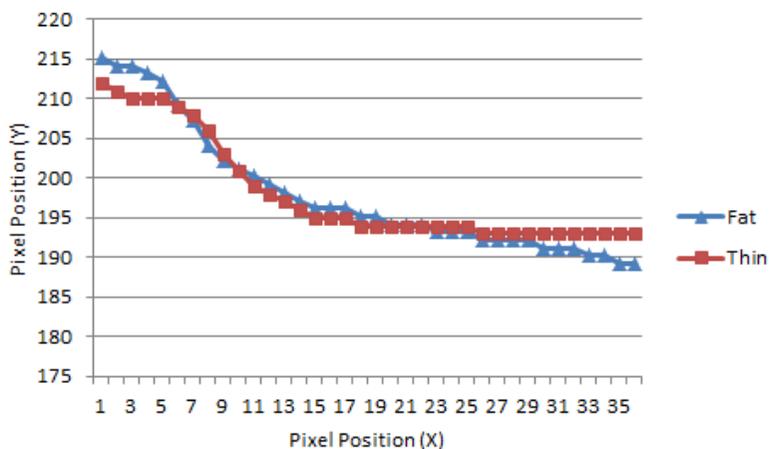


Figure 20. Comparison of thin-fat shoulders with different start points of curves

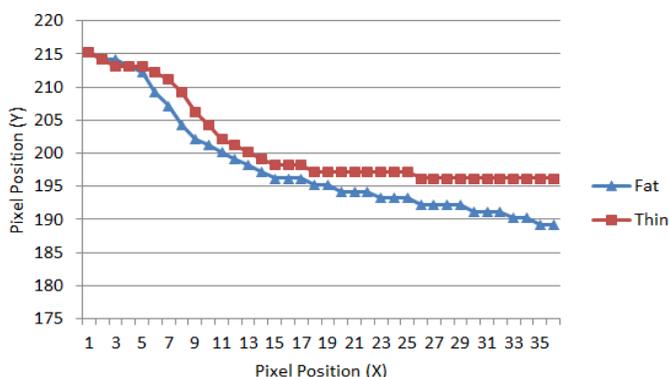


Figure 21. Comparison of thin-fat shoulders with similar start points of curves

Figure 20 and Figure 21 show different start positions in calculating curves similarities between thin and fat shoulder contours. In Figure 20, both contours are shifted with the mean difference value to get largest R^2 value 0.94, which are not necessarily high to show difference between thin and fat shoulder curves, and yields Standard Deviation 1.897. In Figure 21, both contours start from the same start point and yields smaller R^2 value 0.77, which shows better difference in discriminating thin and fat shoulder curve, but it still shows similar Standard Deviation 1.897. For fat shoulders with larger full body width-height ratio, we should use the same start points because it does not change the Standard Deviation value, but shows capability to discriminate better between thin and fat shoulder curves. This should be considered first before combining body and characteristics in human detail features recognition systems.

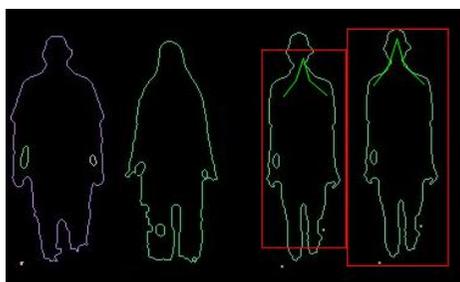


Figure 22. Tracking results of moving human posture with 100% correct thin body posture identification

Figure 22 shows a result from an experiment using HOG-SVM with simulated movement of translation, zoom in/out, variation of illumination of body posture contour figures. Two thin figures in the right side are compared with fat and headscarf figures in the left side. During recognition process, when the figures are detected with the red bounding boxes, they only point two thin figures in the right side and never point the other two figures in the left side, which are 100% correctness of classification (true positive), and zero wrong classification (false positive). This is because the contouring process makes very specific samples taken for HOG-SVM training. Therefore, the machine is able to yield very accurate true positive and zero false positive in testing with the other figures.

5. Conclusions

In this work, we developed further details of moving human visual recognition, which is useful for human tracking, using common characteristics of face and body postures. The more detail features of the object in recognition systems, the higher accuracy will be the results. The learning subsystem must also be able to capture exact patterns as samples of the features without noise. Many samples data should be collected for every possibility of object appearances. Combination of many samples collected on many detail features will complete comparison process to testing object. In application, we demonstrate our method can be useful for visual object classification. Probabilities of personal identification can be increased by using different poses and characteristics of smaller detail features through body postures and face areas. More detail features will richen comparison data samples for higher recognition accuracy.

6. Acknowledgment

The author wish to thank The Indonesian Ministry of Research, Technology, and Higher Education and The Directorate General for Strengthening of Research and Development in supporting this work through Indonesian Institute of Sciences and Institut Teknologi Bandung research and education programs.

7. References

- [1]. T. Ahonen, A. Hadid, M. Pietikainen, "Face Recognition with Local Binary Patterns," *Computer Vision – ECCV*, pp. 469–481, 2004.
- [2]. J. Galbally, S. Marcel, J. Fierrez, "Biometric Antispoofing Methods: A Survey in Face Recognition," *IEEE Access*, vol. 2, pp. 1530-1552, 2014.
- [3]. V. M. Patel, J. K. Pillai, R. Chellapa, "Image and Video-Based Biometrics", T.B. Moeslund et al. (eds.), *Visual Analysis of Humans*, Springer-Verlag London, pp. 437-454, 2011
- [4]. V. Kazemi, J. Sullivan, "One Millisecond Face Alignment with an Ensemble of Regression Trees," *CVPR '14 Proc. of the 2014 IEEE Conf. on Computer Vision and Pattern Recognition*, 2014, pp. 1867-1874.
- [5]. Kinect (2017), <https://developer.microsoft.com/en-us/windows/kinect>
- [6]. S. Zafeiriou, C. Zhang, Z. Zhang, "A survey on face detection in the wild: Past, present and future," *Comput. Vis. Image Underst.*, vol. 138, pp. 1–24, March 2015.
- [7]. L. Antanas, P. Frasconi, F. Costa, T. Tuytelaars, L. De Raedt, "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*, 2012, pp. 171-180.
- [8]. L. Antanas, M. van Otterlo, J. O. M., T. Tuytelaars, L. De Raedt, "A Relational Distance-based Framework for Hierarchical Image Understanding," *Int. Conf. on Pattern Recognition Applications and Methods*, vol. 2, 2012, pp. 206–218.
- [9]. R. T. Olszewski, R. Maxion, D. Siewiorek, D. Banks, "Generalized Feature Extraction for Structural Pattern Recognition in Time-Series Data," Pittsburgh PA: Carnegie Mellon University, February 2001.

- [10]. I. Supriana, Y. Pratama, "Face Recognition New Approach Based on Gradation Contour of Face Color," *Int. Journal on Electrical Engineering and Informatics*, vol. 9, no. 1, pp. 125-138, March 2017.
- [11]. H. T. Chen, Y. Z. He, C. C. Hsu, C. L. Chou, S.Y. Lee, B. S. P. Lin, "Yoga Posture Recognition for Self-training," *Int. Conf. on Multimedia Modeling (MMM 2014)*, 2014, pp. 496-505.
- [12]. G.E. Martinez, O. Mendoza, J.R. Castro, A. Rodriguez-Diaz, "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*, 2015, pp. 1-6.
- [13]. O. Mendoza, P. Melin, G. Licea, "A Hybrid Approach for Image Recognition Combining Type-2 Fuzzy Logic, Modular Neural Networks and the Sugeno Integral," *Inf. Sci. (Ny)*, vol. 179, no. 13, pp. 2078-2101, 2009.
- [14]. C. Torres-Huitzil, "Fast Hardware Architecture for Grey-Level Image Morphology with Flat Structuring Elements," *IET Image Process.*, vol. 8, no. 2, pp. 112-121, 2014.
- [15]. H. Bay, T. Tuytelaars, L. Van Gool, "Surf: Speeded Up Robust Features." *Computer Vision-ECCV*, Springer Berlin Heidelberg, pp.404-417, 2006.
- [16]. S. Leutenegger, M. Chli, R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints," *Computer Vision (ICCV), IEEE International Conference on. IEEE*, 2011.
- [17]. A. Alahi, R. Ortiz, P. Vandergheynst, "Freak: Fast retina keypoint," *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on. IEEE*, 2012.
- [18]. M. Calonder, V. Lepetit, C. Strecha, P. Fua, "Brief: Binary robust independent elementary features." *Computer Vision-ECCV, Springer Berlin Heidelberg*, pp. 778-792, 2010.
- [19]. E. Rublee, V. Rabaud, K. Konolige, G. R. Bradski, "ORB: an Efficient Alternative to SIFT or SURF," *Computer Vision (ICCV), IEEE International Conference on. IEEE*, 2011.
- [20]. D. Andriana, "Linear Function and Inverse Function with Weight Ratio for Improving Learning Speed of Multi-Layer Perceptrons Feed-Forward Neural Networks," *Proc. Int. Conf. on Computer, Control, Informatics and its Applications*, 2013, pp 255-259.
- [21]. D. Andriana, "Contiguous Uniform Deviation for Artificial Neural Network Pattern Recognition," *Advanced Science Lett.* vol. 21, no. 2, pp. 189-191, 2015.
- [22]. D. Andriana, C. Machbub, A. S. Prihatmanto, "Opponent Zigzag Movement Model Capture and Prediction in Robotic Soccer," *International Conf. on Interactive Digital Media*, Bandung, 2015.
- [23]. D. Andriana, A. S. Prihatmanto, E. M. I. Hidayat, I. Supriana, C. Machbub, "Contiguous Uniform Deviation for Multiple Linear Regression in Pattern Recognition," *Journal of Physics, Conference Series*, vol.801, no.1, 2017.
- [24]. Opencv (2017), <http://opencv.org>
- [25]. Dlib C++ Library (2017), <http://dlib.net>
- [26]. U.Toseeb, E. J. Bryant, D. R. T. Keeble, "The Muslim Headscarf and Face Perception: They All Look the Same, Don't They?," *PLoS ONE*, 9, 2, e84754, 2014.
- [27]. R. Yang, Z. Wang, P.-A. Heng, K.-S. Leung, "Fuzzified Choquet Integral with a Fuzzy-valued Integrand and its application on temperature prediction," *IEEE Trans. Syst. Man. Cybern. B. Cybern.*, vol. 38, no. 2, pp. 367-80, 2008.
- [28]. A. F. Tehrani, M. Strickert, E. Hüllermeier, "The Choquet Kernel for Monotone Data," *Esann*, April, pp. 23-25, 2014.
- [29]. The Database of Faces. AT&T Laboratories Cambridge (2017), <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>
- [30]. A. C. Sobieranski, V. F. Chiarella, E. B. Alexandre, R. T. F. Linhares, E. Comunello, A. von Wangenheim, "Color Skin Segmentation Based on Non-linear Distance Metrics," *Progress in Pattern Recog., Image Analysis, Comp. Vis., and App.: Proc. CIARP, (Mexico: Springer) eds. EB-Corrochano and E Hancock*, pp. 143-150, 2014.



Dian Andriana 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 Muhammad Idris 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.