# A Bio-inspired Event-based Size and Position Invariant Human Posture Recognition Algorithm

Shoushun Chen, Berin Martini and Eugenio Culurciello
Electrical Engineering Department, Yale University

***Abstract*—This paper proposes a new approach to recognize human postures in realtime video sequences. The algorithm employs temporal difference imaging between video sequences as input and then decompose the contour of the active object into vectorial line segments. A scheme based on simplified Line Segment Hausdorff Distance combined with projection histograms is proposed to achieve size and position invariance recognition. Consistent with the hierarchical model of the human visual system, sub-sampling techniques are used to represent the object by line segments at multiple resolution levels. The whole classification is described as a coarse to fine procedure. An average realtime recognition rate of 88% is achieved in the experiment. Compared to conventional convolution method, the proposed algorithm reduces the computation cycles by $10-100$ times. This work sets the foundation for size and position invariant object recognition for the implementation of event-based vision systems**

## I. Introduction

Human posture recognition is gaining increasing attention due to its promising application in the area of personal health care, environmental awareness, intelligent visual human machine interface (such as video game systems and human-robot interaction), to name a few. Based on commercially available image sensors and powerful personal computers, impressive research work has been reported for a variety of applications. Works on human gait recognition [1], standing posture recognition with different arm poses [2] and dynamic gesture such as sign language recognition [3] were presented. In general, those approaches first detect moving objects by the analysis of video stream, then extract human silhouettes using background subtraction technique [4][5]. Blob metrics are represented into multiple appearance models [6] and finally posture profiling is conducted based on frame-by-frame posture classification algorithms. Due to the complexity, these algorithms are implemented on powerful computers, even when recognizing only a small subset of human body postures, such as standing, bending, sitting and lying [7]. This will limit the use of these algorithms in real life applications. On the other hand, small and lightweight wireless platforms, such as ultra-mobile PCs or smart cellular phones, are becoming an ubiquitous computation platform. Unfortunately, these devices are still unable to perform power and computation hungry object recognition tasks. In fact, there is a growing gap between the latest computer-based vision algorithms and what is actually implementable in low-complexity hardware.

In addition to the complexity of the above mentioned algorithms, conventional frame based image sensors employed in these systems also lower the energy efficiency. Indeed, the output of conventional image sensors, as a matrix of pixel color values, contain a very high level of redundancy. Large amounts of unimportant data have to be read and processed before obtaining the features of interest [8]. As a matter of fact, the first step of many computer vision algorithms is to remove the background and extract object edges or motion contours [1][9].

In this paper, we present an energy-efficient algorithms for posture recognition. Combined with bio-inspired event-based temporal difference image sensor, which removes the background information without any post processing, the algorithm is able to recognize objects independently of their size and position in the image. The algorithm first stores the address events from a moving object then decompose the contour into vectorial line segments. A hierarchical Hausdorff distance scheme is then employed to measure the similarity of the input line segments with those of a set of library objects. Size and position invariance is achieved by using projection histograms. The proposed approach is innovative due to its high data-encoding efficiency, large saving in computation complexity as well as in its novel way to achieve size and position invariance. The rest of the paper is organized as follows: Section II introduces the system overview. Section III describes the proposed edge feature extraction scheme and the size and position invariant recognition algorithm. Section VI discusses the computation complexity. Section V reports the realtime experimental results. Section VI concludes this paper.

## II. System overview

The architecture of the proposed system is illustrated in Fig.1. It includes an image sensor working at temporal difference mode (MotoTrigger [10]), a hierarchical edge feature extraction unit and a classifier with a set of library postures.

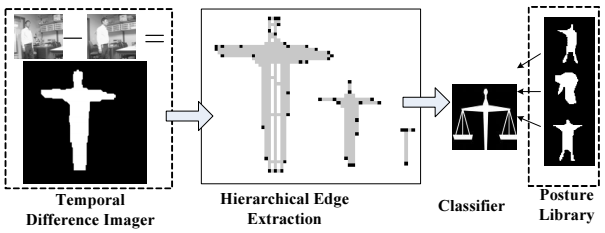



Fig. 1. Block diagram of the system

The temporal difference image sensor compares two continuous image frames and only outputs addresses of those pixels with illumination variance larger than certain threshold. If the scene illumination and object reflectance are constant then the changes in scene reflectance only result from object or viewer movement. Therefore the background information is naturally filtered since the received pixels only come from the active object of interest. This shows great computational efficiency as compared to conventional image sensors used in other systems [1][9]. With the address of the events, an edge feature extraction unit will reorganize the contour of the objects into vectorial line segments. The extracted line segments are fed to a modified Hausdorff distance scheme to measure the similarity of the input line segments with those of a set of library objects. The proposed classifier is able to perform size and position invariance recognition.

## III. Size and position invariant recognition algorithm

The proposed recognition algorithm works in two phases. First the received address events are stored in memory and line segments extraction is performed. Size and position information of the object is also derived in this stage. Inspired by the pioneering work in the modeling of the human visual system [11], which involves tuning and selection of maximal responses at different topological scales, we perform edge extraction at 3 resolution levels. Second, the coordinates of the staring point and ending point of the line segments are sent to the recognition engine for classification.

### A. Hierarchical Body shape extraction

The body shape extraction is achieved in three different stages. First a binary image is built where a "1" is set by decoding the address of the received spike from the image sensor. After that, the image is scanned at four directions, namely horizontal, vertical, $45^o$ and $135^o$. A line segment can be identified by looking at the output of the scanner. For instance, a transition from "0" to "1" denotes a starting point of a line segment while a transition from "1" to "0" indicates an ending point. However, to suppress the internal lines in a thick object, as shown in Fig.2(a), special extraction rule is required. We propose to filter the redundant lines by a special extraction rule as the following statement, "*Ignore a line if both its starting point and ending point are included in another line*". Fig.2 shows the extraction result of a set complicated artificial objects and a human posture. One can notice that, high encoding efficiency is obtained by representing complex objects by limited pairs of dots.

At the third stage, as shown in Fig.1, 2 rounds of sub-sampling are performed on the binary image, each followed by a new scanning procedure. Interestingly, the sub-sampling is implemented by changing the scanner's incremental value instead of physically manipulating the memory content. For instance, to sub-sampling by a factor of two, each time, the column and row address counter just needs to increase by "2". By doing sub-sampling, the object is described as a coarse to fine representation and thus the complexity of the recognition algorithm can be greatly reduced.

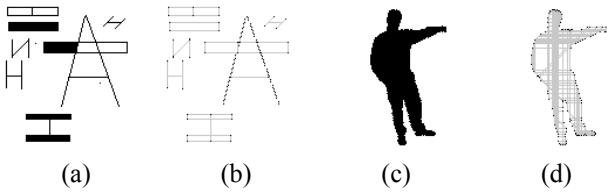



Fig. 2. Edge feature extraction examples for a set of artificial objects and a human posture. (a) and (c) are original images to be extracted. (b) and (d) are reconstructed image using extracted line segments. Each line's two terminals are highlighted by black dots while the internal pixels are in grey color.

### B. Simplified Line Segment Hausdorff Distance

In computer vision, a large number of object recognition algorithms were reported based on Line Segment Hausdorff Distance [12]. For use in our application, we propose the distance of two line segments, as shown in Fig.3, from Line A to Line B, defined as below:

$$D(A \rightarrow B) = P_\theta \times (d_{//1} + d_{//2} + L_A d_\perp) \quad (1)$$

Where $P_\theta$ is an intersection angle penalty coefficient, $d_{//1}$ and $d_{//2}$ is the parallel distance between the two lines' terminals, $L_A$ is the length of line A, $d_\perp$ is the perpendicular distance.

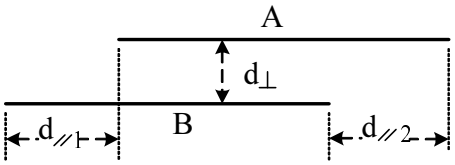



Fig. 3. Displacement measurement of two line segments

Our definition differs with that described in [12], where the parallel distance of two lines is considered to be zero if one line is within the vicinity of the other. That definition improves the matching efficiency for broken lines due to segmentation errors. However, it also causes misjudgements in our application where one posture is part of another distinct one. We take into account the parallel displacements at both terminals, but only weighting the perpendicular distance by the length of the line.

The intersection penalty coefficient also plays an important role. We give it a value of *1* for two lines in parallel and $\infty$ for two lines in perpendicular. For two lines intersecting at a angle of $45^o$ or $135^o$, we investigated the effect of this coefficient to the distance measurement of postures by Matlab simulation. As shown in Fig. 4, one can see that the selectivity increases with the coefficient until saturated at certain point, which implies an optimal value.

### C. Size and position invariance recognition

Usually two objects needs to be aligned before comparing their distance. As witnessed in face recognition, two face

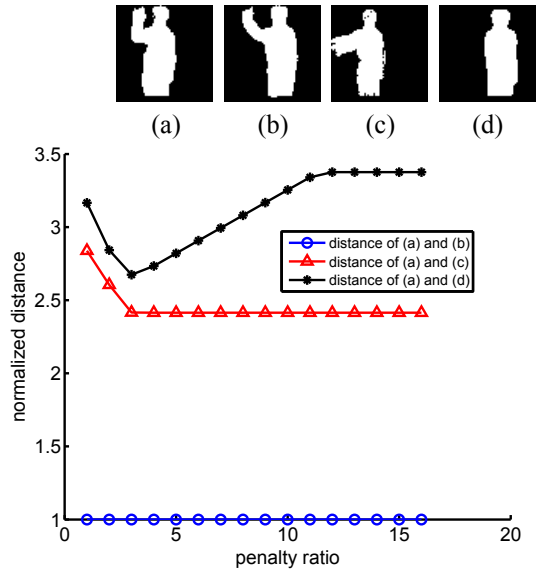



Fig. 4. Distance measurement of several postures under different 45º intersection distance penalty. (a) and (b) belongs to the same type of posture and their distance is used to normalize the distance of (a)and (c), (a) and (d).

images are aligned based on the location of the eyes [12]. To achieve this for human postures or even generic objects, we propose to align two objects using their center position. In the line segment extraction unit, we build both row and column histogram counters. Each time a pixel's address is received, by decoding its row and column address, the corresponding counter will increase by one.

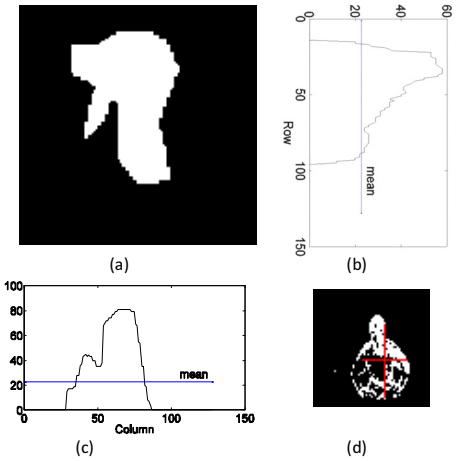



Fig. 5. The Projection Histograms of a posture. (b) and (c) shows the row and column Projection Histogram, respectively. (d) shows the size and center information of another two postures.

As depicted in Fig.5, the row and column histogram can directly reflect the position of an object. We define the mean value of the histogram as a threshold, to remove the effects of noise. The first row address $x1$ and the last row address $x2$ where histogram value exceeds the threshold can be obtained and their center is considered as the row center of the object. Column center can be obtained using the same technique. Interestingly, projection histograms not only provides position information, it also reports the size of the object. We propose to estimate the horizontal size of the object by the distance between $x1$ and $x2$. For the example posture shown in Fig.5(d), the cross shows the center and "X-Y" size information found using the above technique.

With the center and size information, both position and size invariant recognition can be achieved. This is done by first aligning the center of the input object and dictionary object and followed by a resizing operation to make them have the same size. This operation only involves a subtraction and a multiplication operation applied to the coordinates of the input line segments.

## IV. Implementation Complexity Analysis

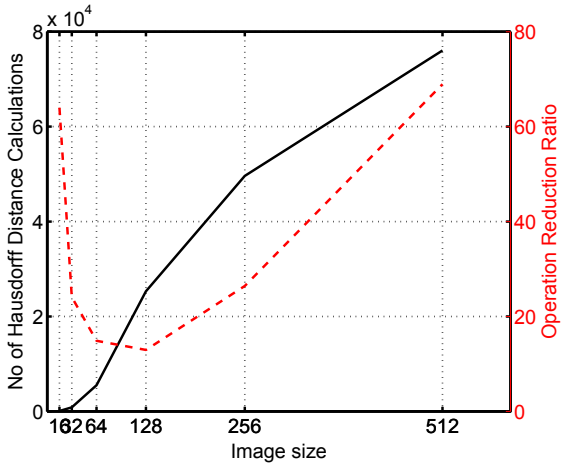



Fig. 6. Computation saving using our proposed algorithm compared to conventional convolution method. It assumes that the latter method has already achieved size and position invariance, which needs additional computations. Even this, our method shows a reduction up to 100 times of operation cycle.

Compared to the standard approach, the proposed algorithm achieves great computational saving, resulting from several novel techniques. First, the object of interest is directly obtained from the output of temporal difference image sensor without any additional processing. Secondly, the contour of the object is decomposed into limited number of line segments. In fact, each line segments distance calculation involves only 3 multiplication and 2 summation operations. Finally the use of hierarchically coarse to fine classification scheme further reduces the number of dictionary candidates. Fig.6 shows the calculated operation reduction factor when our algorithm is compared to standard object-recognition approaches used in computer vision. Even for small imaging size we can reduce the computation by $10 - 100\times$ when compared to classic convolution techniques. And this a conservative number because it assumes that size and position invariance is already computed by some other means, while they are an integral part of our approach.

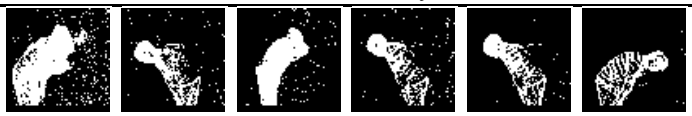
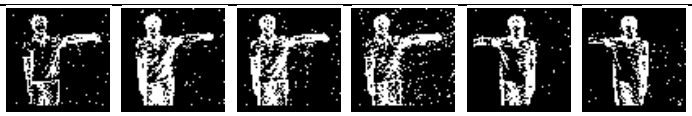
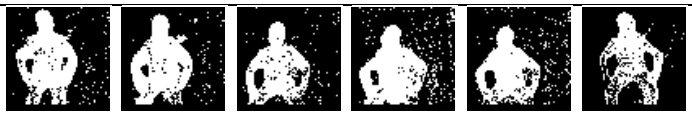
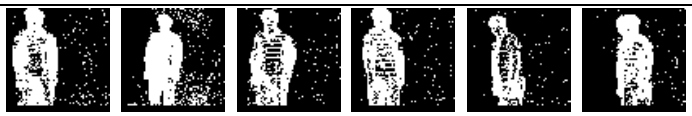
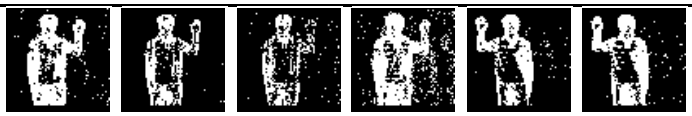
| Posture Group | Postures Library | Success Rate |
|---|---|---|
| bend |  | 192/200=96% |
| raise-1-hand |  | 152/200=76% |
| raise-2-hand |  | 189/200=94% |
| squat |  | 178/200=89% |
| stand |  | 190/200=95% |
| swing hand |  | 166/200=83% |

Fig. 7. Six groups of postures used for experiment. For each group, 6 postures are used as library units while a larger set are used as test entries. It can be noticed that, our algorithm can successfully distinguish quite similar postures, such as "raise-1-hand" and "swinghand". Due to the limitation of operational environment, we did not try experiment on "lying down" postures.

## V. Experimental Results

The proposed algorithm is implemented to perform realtime posture recognition. The dictionary are some postures available in laboratory environment, namely "bending", "standing", etc, as shown in Fig.7. To evaluate the exact figures of successful recognition, we take a large set of snap shots of postures and put several of them (6 for each group of postures) as dictionary objects and the rests are used as test data. As shown in the rightmost column in Fig.7, good accuracy up to $97\%$ is achieved. We can even distinguish between quite similar postures, for example, "raise-1-hand" and "swing hand". The standing postures show a relative low recognition rate. This is explained by the fact that these postures sometimes appear as a part of another posture, on which our definition of line segment distance (Eq.1) is to address.

## VI. Conclusion

This paper reports a size and position invariant human posture recognition algorithm. The image is first acquired using an address event temporal difference image sensor and followed by a bio-inspired hierarchical line segment extraction unit. A simplified line segment Hausdorff distance scheme is employed for similarity measurement while size and position invariance is achieved by deriving size and position information from Projection Histograms. The proposed algorithm achieves 88% average recognition rate while features $10-100\times$ computational saving as compared to conventional approach.

## VII. Acknowledgements

This project was funded in part by NSF award 0622133.

## References

[1] Han Su and Feng-Gang Huang, "Human gait recognition based on motion analysis," 2005 International Conference on Machine Learning and Cybernetics, Aug. 2005, vol. 7, pp. 4464-468.

[2] Boulay B., et *al.*, "Posture recognition with a 3D human model," The IEE International Symposium on Imaging for Crime Detection and Prevention, Jun. 2005, pp.135-138.

[3] Isaacs J. and Foo S., "Hand pose estimation for American sign language recognition," 36th Southeastern Symposium on System Theory, 2004, pp. 132-136.

[4] Takahashi K., et *al.*,, "Remarks on Real-Time Human Posture Estimation from Silhouette Image Using Neural Network," IEEE International Conference on Systems, Man and Cybernetics, Oct. 2004, pp. 370-375.

[5] E. H-Jaraha, et *al.*, "Detected motion classification with a double-background and a Neighborhood-based difference," Pattern Recognition Letters, 2003, Vol. 24, pp. 2079-2092.

[6] L.H.W. Aloysius, et *al.*, "Human posture recognition in video sequence using Pseudo 2-D Hidden Markov Models," 8th Control, Automation, Robotics and Vision Conference, 2004, Dec. 2004, pp. 712-716.

[7] Spagnolo P., et *al.*, "Posture estimation in visual surveillance of archaeological sites," IEEE Conference on Advanced Video and Signal Based Surveillance, Jul. 2003, pp. 277-283.

[8] E. Culurciello and A. Savvides, "Address-Event Image Sensor Network,ISCAS 2006, 21-24 May 2006, pp. 955-958.

[9] Triesch J. and von der Malsburg C.,"A system for person-independent hand posture recognition against complex backgrounds," IEEE IEEE TPAMI, Vol. 23, Dec. 2001, pp. 1449-1453.

[10] Z.M. Fu, E. Culurciello, "A 1.2mW CMOS Temporal-Difference Image Sensor for Sensor Networks," ISCAS 2008, May 2008, pp. 1064-1067.

[11] Thomas Serre, "Learning a Dictionary of Shape-Components in Visual Cortex: Comparison with Neurons, Humans and Machines," Ph.D Thesis of Brain and Cognitive Sciences Department, MIT, April, 2006.

[12] Yongsheng Gao and Leung M.K.H., "Face recognition using line edge map," IEEE TPAMI, Jun. 2002, pp. 764-779.