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

Shoushun Chen[1], Fopefolu Folowosele[2], Dongsoo Kim[1], R. Jacob Vogelstein[2], Ralph Etienne-Cummings[2] and Eugenio Culurciello[1]
[1] Electrical Engineering Department, Yale University
[2] Dept. of Electrical & Computer Engineering, Johns Hopkins University

***Abstract*— In this paper we report a size and position invariant human posture recognition algorithm. The algorithm employs a simplified line segment Hausdorff distance classification and uses projection histograms to achieve size and position invariance. Compared to other existing method utilizing line segment Hausdorff distance, the proposed algorithm reduces the computation complexity by 36000 times, for our test images. Combining bio-inspired event-based image acquisition and hardware friendly feature extraction and classification algorithm will lead to a promising technology for use in wireless sensor network.**

I. Introduction

Recent advancement in wireless sensor networks have provided new opportunities to explore low-power sensory systems and their applications. In a recent review [1], a number of promising current and future wireless sensor networks technologies, for home health monitoring, were listed and discussed. Such systems could form the basis for a future "smart" home in which an ambient awareness of the home's occupants is maintained through an ecosystem of ubiquitous connectivity, disappearing devices, highly available services, and multi-modal sensing. In such systems, posture recognition could play an important part, however, the current implementation of posture recognition tends to be large and complex. Even when recognizing only a small subset of human body postures, such as standing, bending, sitting and lying, a complete computer system is needed [2]. Besides the complex and sequential nature of the algorithms, the way image or video data is collected makes current systems unsuitable for use in a low bandwidth and low power environment. Advanced deep sub-micron technologies have enabled higher resolution and higher frame rate image sensors featuring improved image and video quality but at the expense of increased output bandwidth. Due to power considerations communication links among wireless sensor nodes are often low bandwidth protocols, such as ZigBee (up to 250 kbps) and Bluetooth (up to 1 Mbps). Even at the data rate of Bluetooth, conventional image sensor can barely stream an uncompressed 320×240 binary video at 13 frame/s. To avoid communication of raw data over wireless channels, on-chip pre-processing mechanisms are utilized to filter out redundant information and extract items of interest. Image compression alleviates some of the bandwidth burden but unfortunately at the cost of additional complex processing which translate into high power consumption. This inherent bandwidth problem stems from the fact that these frame-based cameras store and process image information in a matrix form. This format is simple and practical for image storage, but it is not ideal for image processing and feature-extraction. Another drawback of using intensity-based images is that they contain a very high level of data redundancy, which, while useful for human interpretation and retrieval, is not ideal for machine-based processing.

In this paper, we present a hardware-oriented size and position invariant human posture recognition algorithm. The algorithm is implemented as a layer in a biologically-plausible hierarchical model of visual information system [3]. In previous work we have demonstrated the implementation of the first stages of the model (from the retina to simple cells) [4][5]. We now propose a novel contour extraction scheme, organizing spars simple cells into line-segments which are used in a recognition algorithm based on a hierarchical Hausdorff distance. 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 a novel way to achieve size and position invariance. The remainder of the paper is organized as follows: Section II introduces the system overview. Section III describes the bio-inspired event based image acquisition and filtering system. Section IV presents the proposed edge feature extraction scheme and the size and position invariant recognition algorithm. Section V reports the simulation results on a set of real images and Section VI discusses the computation complexity. Section VII concludes this paper.

II. System Overview

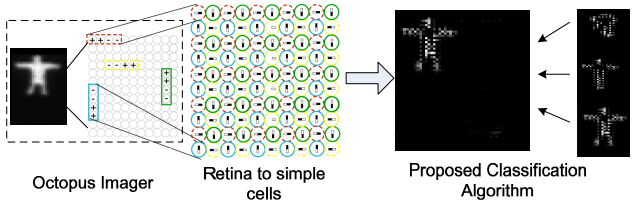


Fig. 1. Block diagram of the system based on the biologically-plausible hierarchical model of visual information.

As shown in Fig.1, the proposed system consists of an address event image sensor [6] and the integrate-and-fire array transceiver (IFAT) system [4][5] to obtain simple cells, followed by the proposed object recognition processor. The image

sensor consists of 80×60 pixels. Unlike in frame based image sensors where pixel's current or voltage is queried for information after a fixed time interval, the pixels push information to their receiver once they meet certain threshold condition. Therefore, this type of image sensor does not produce "image frames" but rather a biologically-inspired stream of events. The address events are then processed by a simple cell network implemented on the IFAT. Simple cells are oriented spatial filters that detect local changes in contrast. Each simple cell receives inputs from four consecutive photodetectors. After simple cell extraction, we perform line segments extraction to further compress the data by organizing the spare cells into vectorial line segments. Recognition is then conducted by comparing the input object and a set of dictionary objects based on a simplified line segment Hausdorff distance scheme while size and position invariance is achieved by deriving size and position information from the projection histograms.

III. Image Acquisition and Filtering System

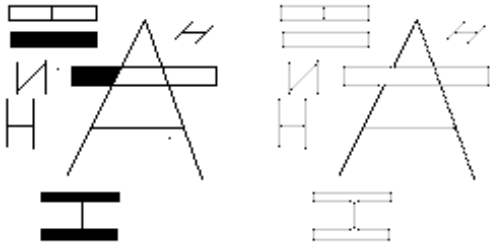
The integrate-and-fire array transceiver (IFAT) system is based on a reconfigurable silicon array of general-purpose integrate-and-fire (IF) neurons for processing spike-based sensory information in real time [4][5]. The system consists of up to four IF chips (with a total of 9600 neurons), digital memory (RAM), a digital to analog converter (DAC) and an FPGA. The IFAT acts as an address-event transceiver, communicating incoming and outgoing events over an asynchronous bus.

The external source providing spikes to the IFAT system is the octopus retina. The silicon octopus retina is unlike its biological counterpart, in that spike outputs are collected on an asynchronous bus and transmitted serially off-chip (in contrast to the dedicated axon along which each photosensor's output travels to its targets allowing for parallel connectivity in biology). It consists of a 60×80 array of integrate-and-fire neurons that translates light intensity levels into inter-spike interval times at each pixel. When the FPGA receives a request from the octopus retina, it reads the address bits and stores it as the presynaptic neuron address. It then uses this address as an index into the RAM to obtain a set of postsynaptic neuron addresses and the synaptic parameters associated with them. The synaptic weights are set up on the IF chips and the analog synaptic equilibrium potentials are established by the DAC. The events are then serialized as they are sent to the chip. Output spikes are re-routed to the FPGA to allow for recurrent connections. Each simple (IFAT) cell receives inputs from four consecutive octopus retina neurons, two of which make excitatory synapses while the other two make inhibitory synapses. The excitatory and inhibitory synaptic weights are balanced so that the simple cells do not respond to uniform light. We implemented four types of simple cells. One type was oriented in the vertical direction and responded to dark-to-light transitions and a second type responded to light-to-dark transitions (Fig.1). The two other simple cell types were oriented in the horizontal direction and responded to light-to-dark and dark-to-light transitions respectively. The IFAT spikes corresponding to the simple cell outputs for the posture images are collected to feed the recognition algorithm.

IV. Recognition Algorithm based on Simplified Line Segment Hausdorff Distance

A. Edge Feature Extraction

The received spikes are stored in memory and the edges defining the contour of the object are extracted for classification. This approach is inspired by works reported in computer vision [7][8], where algorithms based on Line Edge Maps (LEM) were proposed and proved to be a viable way for face recognition. As our application is classifying only basic human postures such as standing, bending, squatting and lying, a reduced set of line segments can be used. Lines are extracted in four orientations, horizontal (0^o) line, vertical (90^o) line and two diagonals ($45^o/135^o$).



(a) (b)

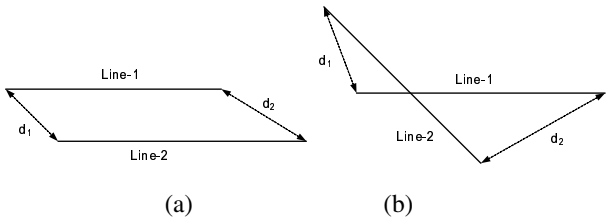
Fig. 2. Line segments extraction on a complicated figure composed by objects with different orientation and thickness. (a) shows the original image (b) shows the extraction results. Each pair of highlighted dots denote an extracted line's two terminals.

Fig.2 shows the extraction result of a complicated figure composed of objects with different orientation and thickness. Each pair of highlighted dots denotes two terminals of an extracted line. High encoding efficiency is obtained by presenting the objects by pairs of dots. It's important to note that, only the outline of the thick objects are permitted to be extracted while the internal lines are filtered. This is achieved 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*".

B. Simplified Line Segment Hausdorff Distance

Hausdorff distance is originally defined for distance measurement between two point sets. Unlike most shape comparison methods that build a one-to-one correspondence between a model and a test image, the Hausdorff distance can be calculated without explicit pairing of points. The Hausdorff distance was further extended to measure the displacement between lines [7]. The distance between two line segments was measured by combining orientation distance, parallel distance and perpendicular distance. In our research, we only consider 45^o angles and so, measurements are further simplified as below:

- For two parallel line segments, distance is defined as the sum of the distance between the two lines' terminals.
- For two perpendicular line segments, distance is defined as infinity.
- For two intersecting lines at 45^o angle, distance is defined as α times of the sum of the distance between the two lines' terminals.



(a) (b)

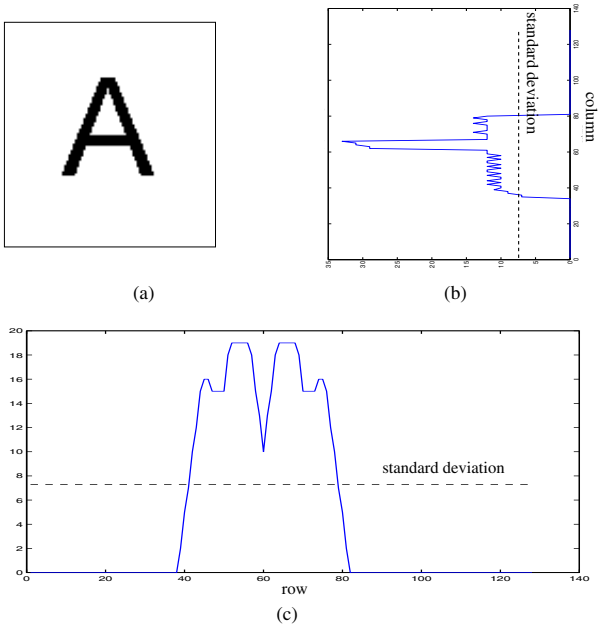
Fig. 3. Displacement measurement of two lines in parallel (a) and 45^o intersecting angle(b)

As shown in Fig.3, the distance between the two lines' terminals is denoted as $d1$ and $d2$ respectively. For the case of Fig.3(a), the two lines' distance is represented by $(d1 + d2)$ while $\alpha \times (d1 + d2)$ for the case of Fig.3(b). To filter out the effect of noise lines, the displacement measurement is further weighted by the lines' length. The final mismatch measurement from a lines set M^l to another set T^l is defined as:

$$h(M^l, T^l) = \frac{1}{\sum_{m_i^l \in M^l} l_{m_i^l}} \sum_{m_i^l \in M^l} l_{m_i^l} \cdot \min_{t_j^l \in T^l} d(m_i^l, t_j^l) \quad (1)$$

where $l_{m_i^l}$ is the length of a line segment m_i^l.

C. Position Invariance Recognition



(a) (b) (c)

Fig. 4. The Projection Histograms of a letter "A". (b) and (c) shows the row and column Projection Histogram, respectively.

In face recognition, to achieve position invariance, two face images are aligned based on the location of the eyes [7]. To achieve alignment between generic objects, we propose to align two objects using their center position as derived from their Row and Column Projection Histograms. Thanks to the data format out of the image sensor, Projection Histograms can be easily obtained by assigning a counter for each row and column to count the number of events. Each time a pixel's address is received, by decoding its row and column address, the corresponding counter will increase by one. As depicted in Fig.4, the row and column histogram directly reflects the position of the object. To remove the effects of noise, standard deviation is used as a threshold. 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.

D. Size Invariance Recognition

Interestingly, projection histograms not only help to achieve position invariant recognition, but also provide the size information of the object which leads to size invariance recognition. For the example shown in Fig.4, the distance between $x1$ and $x2$ is considered as the size of the object. Then by normalizing the size of the test object and library object, size invariance recognition can be achieved.

Combining both size and position invariance, the Line Segment Hausdorff distance defined by Eq.1 is updated to

$$h(M^l, T^l) = \frac{\sum_{m_i^l \in M^l} l_{m_i^l} \cdot \min_{t_j^l \in T^l} d(\frac{m_i^l - C_{M^l}}{S_{M^l}}, \frac{t_j^l - C_{T^l}}{S_{T^l}})}{\sum_{m_i^l \in M^l} l_{m_i^l}} \quad (2)$$

where S_{T^l} and S_{M^l} is the size of the test and library object, C_{T^l} and C_{M^l} is the center of the test and library object, respectively.

The above described recognition algorithm appears to be a brute force procedure, however, the procedure implemented is in fact performed in a hierarchical fasion. As the object is described by four set of different orientation lines, the number of candidate objects can be narrowed down by calculating Hausdorff distance set by set. For instance, after calculating the distance of horizontal lines, only the first half nearest objects will be permitted for vertical line distance calculation.

V. Simulation Results

Extensive simulations were performed to evaluate the performance of our algorithm on size and position invariant recognition. Fig.5 shows the simulation results on a set of human postures. The top left posture is set as an original reference, then shifted left in the scene (1st row, 2nd column), vertically scaled (1st row, 3rd column), horizontally scaled (2nd row, 1st column), both shifted and horizontally scaled (2nd row, 2nd column). Another similar but different posture is used (2nd row, 3rd column) and the distances between the 1st image to the rest ones are calculated. It's clear that, the shifted and scaled "same" postures have a smaller distance when compared to the "different" one. Therefore, it's shown that size and position invariant recognition has been achieved.

Then we perform recognition on the output data of the simple cells extracted by the IFAT system. As a proof of concept, we use a number of black and white pictures of human posture as test and dictionary data. Image and simple cells data, in the format of address events, is first quantized by a threshold number of events. The quantized address events are used as input for the line segments extraction unit. Fig.6 shows the simulation output. The figures are mainly standing

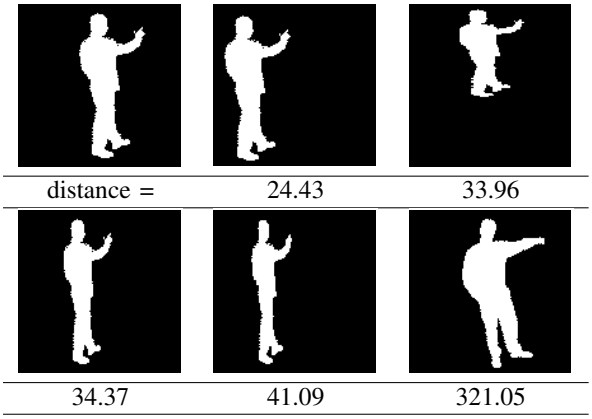


Fig. 5. Simulation results of size and position invariance recognition. The top left image is an original reference, the rest ones except the last figure are obtained by either position shifted, vertically or horizontally scaled. The line segment Hausdorff distance between the first image and reset ones were calculated and shown under each image.

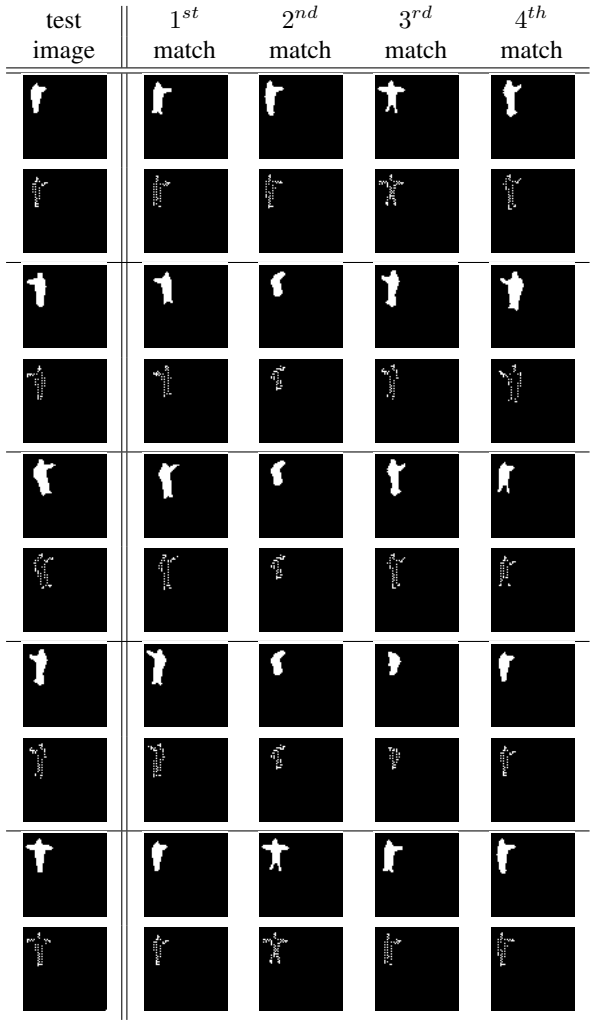


Fig. 6. Simulation results of human posture recognition. The first column is the test image while the right four columns are library images ranked by similarity. For every two rows, the first row shows the raw image acquired by the Octopus imager plus IFAT processor, while the second shows the corresponding S1 layer image.

postures with different leg and arm poses. Each is in turn used as a test (the 1st column) and the reset are used as dictionary. Then the dictionary figures are ranked by similarity, shown in the right four columns.

VI. Implementation Complexity

The proposed algorithm shows great hardware implementation efficiency. Each line segments distance calculation needs 3 multiplication and 2 summation. For the above 16 test figures, the average number of extracted line segments is 15 which implies 4 line segments for each orientation. Then for each figure, by applying the hierarchical comparison approach mentioned in the previous section, the total number of line segments distance to be calculated is $4 \times 4 \times 16 \times (1 + 1/2 + 1/4 + 1/8) \times 2 = 960$. To achieve position invariance (without size invariance), conventional approach needs to try all possible reference positions. For straightforward brute force approach, the total number of Euclidean distance to be calculated is $(60 \times 80)^2 \times 16 = 368640000$. Even using line segments Hausdorff distance proposed in [7], the number of the calculation is still $15 \times 15 \times (60 \times 80) \times 16 \times 2 = 34560000$, 36000 times of our solution.

VII. Conclusion

In this paper, a size and position invariant human posture recognition algorithm is proposed. The similarity measurement is based on a simplified line segment Hausdorff distance scheme while size and position invariance is achieved by deriving size and position information from Projection Histograms. The proposed algorithm is evaluated by performing recognition on the first stages of a biologically-plausible hierarchical model of visual information system. Due to its efficient data encoding and hardware friendly architecture, the feature exaction circuits will be implemented into the next generation event based image sensor while the the recognition algorithm will be implemented on FPGA or ASIC. This combination will lead to promising technology for inclusion in wireless sensor network.

VIII. Acknowledgements

This project was funded by NSF award 0622133 and Peregrine Semiconductors. We also thank Berin Martini for collecting data from Octopus Imager.

References

[1] Baker, Chris R, et *al.*, "Wireless Sensor Networks for Home Health Care," 21st International Conference on Advanced Information Networking and Applications Workshops, May 2007, pp. 832-837.

[2] 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.

[3] M. Riesenhuber and T. Poggio, "Hierarchical models of object recognition in cortex," Nature Neuroscience,1999, vol. 2, no. 11, pp. 1019-1025.

[4] Mallik U., et *al.*, "A real-time spike-domain sensory information processing system," ISCAS 2005, May 2005, pp. 1919-1922.

[5] R.J. Vogelstein, et *al.*, "A multichip neuromorphic system for spike-based visual information processing," Neural Computation, vol. 19, pp. 2281-2300, 2007.

[6] Culurciello E., et *al.*, "A biomorphic digital image sensor, " IEEE JSSC, Vol. 38, No. 2, 2003, pp. 281-294.

[7] Yongsheng Gao and Leung M.K.H., "Face recognition using line edge map," IEEE Transaction on Pattern Analysis and Machine Intelligence, Jun. 2002, pp. 764-779.

[8] Hye-mi Kim, et *al.*, "Face Recognition using 3D Line Edge Map Robust to Expression Changes and Occlusions," The 9th International Conference on Advanced Communication Technology, Feb. 2007, pp. 357-362.