MKinect-A Novel Methodology to Recognize the Composite Human Gesture for Microsoft Kinect

S. Sibi Chakkaravarthy, S. Thiyagarajan, G. Arun Karthick and K.A. Varun Kumar

Abstract *--- A novel method used in Kinect for recognizing a human body in all textural posture was proposed here. The method is to recognize the human gesture so that the human can be identified easily .If the human is subjected to the identification where the intensity of environment lapse, "mapping" is applied to separate the human from environmental background. The system authenticates & actuates by means of several procedural actions such as abstract formation for human body & necessary actions to be taken out to predict the skeleton, point generation based on 3d posture based point value of an image predicted for human body skeleton, segmentation of human body, generating value for each segmentation and generating code for the value obtained from the predicted posture. Experimental results demonstrating the proposed approach was also included.*

Keywords--- *Kinect, Mapping, Threshold Values, Pixel Intensity, Human Postures, Gesture*

I. INTRODUCTION

THE goal of this work is to detect a human figure image and localize his joints and limbs along with their associated pixel masks. In this work we attempt to handle the problem in a general way. The dataset we use is a collection of a human, varying dramatically in pose and clothing. The approach that we take is to use segmentation to guide our recognition algorithm to salient bits of the image. We use this segmentation approach to build limb and torso detectors, the outputs of which are assembled into human figures. We present quantitative results on torso localization; in addition to short listed full body configurations. This paper presents a system that can automatically recognize human body postures in video sequences. The considered postures are standing, sitting, skating, walking, squatting, and lying. The recognition is based on data fusion using the belief theory. The data come from the person's 2D segmentation and from their face localization. It consists in distance measurements relative to a reference posture ("Da Vinci posture": standing, arms stretched horizontally).The segmentation is based on an adaptive background removal algorithm. The face localization process uses skin detection based on color information with an adaptive threshold. The efficiency and the limits of the recognition system are highlighted thanks to the analysis of a great number of results. This system allows real-time processing.

A new method for representing and recognizing human body movements is presented. The basic idea is to identify sets of constraints that are diagnostic of a movement: expressed using body-centered coordinates such as joint angles and in force only during a particular movement. According Lee.W.Campbell,Aaron F.Bobick[12] says that Assuming the availability of Cartesian tracking data, we develop techniques for a representation of movements defined by space curves in subspaces of a "phase space." The phase space has axes of joint angles and torso location and attitude, and the axes of the subspaces are subsets of the axes of the phase space. Using this representation we develop a system for learning new movements from ground truth data by searching for constraints. We then use the learned representation for recognizing movements in unsegmented data. We train and test the system on nine fundamental steps from classical ballet performed by two dancers; the system accurately recognizes the movements in the unsegmented stream of motion.

Approaches to recognizing 3D human body postures from a single image have recently become increasingly popular. While they do not suffer from many of the problems that affect more traditional recursive body tracking techniques, most of them have only been demonstrated in cases where clean body silhouettes can be extracted, for example using background subtraction, which is very restrictive. A key exception is the work reported. Combining a hierarchy of templates and effectively using the chamfer distance has made the approach applicable to more challenging cases such as the one of a moving camera on a car. However, even then, the algorithm tends to produce many false positives, especially when the background is cluttered. As a result, in practice, it is used in conjunction with a stereo rig both to narrow the initial search area and to filter out false detections from the background. We improve upon this approach and achieve very low rates of both false positives and negatives by incorporating motion information into our templates. It lets us differentiate between actual people and static objects whose outlines roughly resemble those of a human, which are surprisingly numerous. As illustrated this is key to avoiding misdetections. This is of course a well known fact and optical flow methods have been proposed to detect moving humans We chose this specific posture both because it is very characteristic and because it could easily be used to initialize a more traditional recursive tracking algorithm to recover the in-between body poses.

S. Sibi Chakkaravarthy, M. Tech Scholars, Vel Tech DR.RR&DR.SR Technical University, Chennai. E-mail: sb.sibi@gmail.com

S. Thiyagarajan, M. Tech Scholars, Vel Tech DR.RR&DR.SR Technical University, Chennai. E-mail: sankarthiyagarajan@gmail.com

G. Arun Karthick, M. Tech Scholars, Vel Tech DR.RR&DR.SR Technical University, Chennai. E-mail: arunkar90@gmail.com

K.A. Varun Kumar, M. Tech Scholars, Vel Tech DR.RR&DR.SR Technical University, Chennai.

As shown in, we obtain good results even when the background is muddled and background detraction is unrealistic because the camera moves. Note that the subjects move closer or further so that their apparent scale changes and turn so that the angle from which they are seen also varies. In this example, no stereo data or information about the ground plane was required to eliminate false-positives. Our method retains its effectiveness indoors, outdoors, and under difficult lighting conditions. Furthermore, because the detected templates are projections of 3D models, we can map them back to full 3D poses. Note that, even though we chose a specific motion to test it, our approach is generic and could be applied to any other actions that all people perform in roughly similar ways but with substantial individual variations. For example, there also are characteristic postures for somebody sitting on a chair or climbing stairs. In the area of sports, we could use a small number of templates to represent the consecutive postures of a tennis player hitting the ball with a forehand, a backhand, or a serve, as is done in. We could similarly handle the transition between the upswing and the downswing for a golfer. In short, characteristic postures are common in human motion and, therefore, worth finding. The only requirement for applying our method is that a representative motion database can be built. In the reminder of the paper we first briefly discuss earlier approaches. We then introduce our approach to body pose detection and present a number of results obtained in challenging conditions. Finally, we discuss possible extensions.

II. Related Work

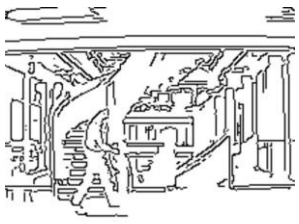
Until recently, most approaches to capturing human 3D motion from video relied on recursive frame-to-frame pose estimation. While effective in some cases, these techniques usually require manual initialization and re-initialization if the tracking fails. Chen Wu and Aghajan, H.[7] says a 3D human body model is employed as the convergence point of spatiotemporal and feature fusion. Cuong Tran , Anup Doshi, Mohan Manubhai Trivedi[25] says a gesture of body recognition is only through the modeling points and clips joint on a body.

As a result, there is now increasing interest for techniques that can detect a 3D body pose from individual frames of a monocular video sequence. One approach is to use classification to detect people in images, but it does not provide either a pose or a precise outline. Furthermore, such global approaches tend to be very occlusion sensitive. Instead of detecting the body as a whole, a different tack is to look for individual body parts and then to try assembling them to retrieve the pose. This can be done by minimizing an appropriate criterion, for example using an A¤ algorithm. This has the potential to retrieve human bodies under arbitrary poses and in the presence of occlusions. Furthermore it can be done in a computationally effective way using pictorial structures. However, it can easily become confused because there are many limb-like objects in real world images. Another class of approaches relies on techniques such as background subtraction to produce silhouettes that can then be analyzed. Several methods learn during an offline stage a mapping between the visual input space formed by the silhouettes and the 3D pose space from examples collected manually or created using graphics software. For example, uses multilayer perceptions to map the silhouette represented by its moments to the 3D pose. In the mapping is performed using robust locally weighted regression over nearest neighbors that are efficiently retrieved using hash tables. In, it is done indirectly via manifolds embedded in low dimensional spaces, where each manifold corresponds to the subset of silhouettes for walking motion seen from a particular viewpoint. Local Linear Embedding is used to map the manifolds to both the silhouettes and the 3D pose. In, the mapping between the couple formed by an extracted silhouette and a predicted pose to the corresponding 3D pose is established using Relevant Vector Machine. While these works introduce powerful tools to associate 3D poses to detected silhouettes, they tend to be of limited practical use because they require relatively clean silhouettes that are not always easy to obtain. A more robust way to match global silhouettes against image contours is to use both a hierarchy of templates and the chamfer distance, an approach originally introduced in and extended in. This produces excellent results when applied to difficult outdoor images. However, it seems to have a relatively high false detection rate. Reducing this rate involves either introducing *a priori* assumptions about where people can be or incorporating additional processing such as texture classification or stereo verification in the context of hand tracking, also relies on the chamfer distance and a tree structure quite similar to the hierarchy of templates for efficiency. In this case, the false positives and negatives problem is avoided by assuming that one and only one hand is present in the image. Bayesian tracking is combined with detection to disambiguate the hand pose.

By contrast to these earlier approaches, our method, which also relies on global silhouettes matching, includes an original way to take motion into account to avoid false positives. Such information was also exploited in for human action recognition, but only under the assumption that preprocessed and centered sub images of the people are available. In our case we directly use the full images as input.

III. Approach

In this section, we describe how we introduce motion information into the point matching process. This is done on the sole basis of the noisy and potentially incomplete generated points over human body that can realistically be extracted from images of cluttered scenes acquired by kinect camera.



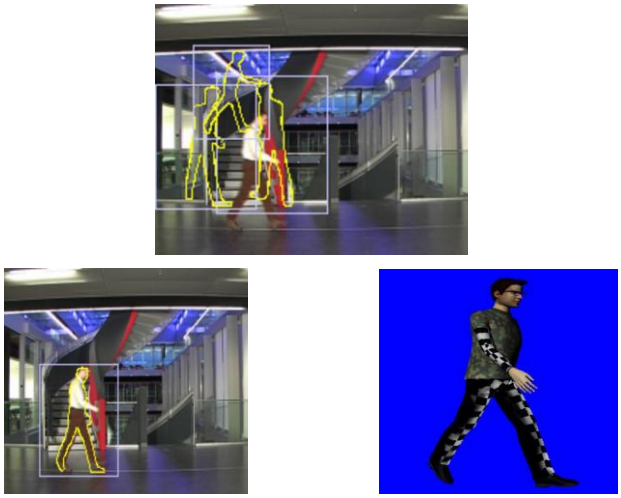


Figure 1: Recognizing a Human Body

IV. Recognition by Generating Points

Figure 1 states that the composite human postures is identified using the 3d biometrical camera or kinect . Here, we focus on the part of the walking cycle where both Feet and legs are on the ground and the angle between them Are the greatest, and use motion capture data and graphics Software to create a database of templates. We first used a optical motion capture system .first the victim image is taken and the appropriate points has been generated for the given image .and the newly generated point is compared with the based image (generated code for the points).if it resides the same. The motion is recognized. Each template is made of a short sequence of that includes a key frame, that is the frame representing the specific walking pose and which is always taken to be the middle frame in the sequence. The image points are represented

As sets of oriented pixels that can be efficiently Matched against image sequences, as will be discussed in practice we use 3 frame (1.abstract formation, 2.generating points for the image and 3.generating code for the points (image value based) sequences.

Figure 3: Abstract formation

For a function $u(x,y,z,t)$ of three spatial variables (x,y,z) and the time variable t, the heat equation is

$$\frac{\partial u}{\partial t} - \alpha\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right) = 0$$

Also written

$$\frac{\partial u}{\partial t} - \alpha\Delta u = 0$$

or alternatively

$$\frac{\partial u}{\partial t} - \alpha\nabla^2 u = 0$$

The Top row corresponds to a profile view in which The ‘I represent the angles between the two legs. Here, we

V. Point Generating

As in previous approaches, we rely on Chamfer distance, efficiently computed using the Distance Transform (DT) of the input image, to match silhouettes to individual input images. However, we have endeavored to increase its robustness.

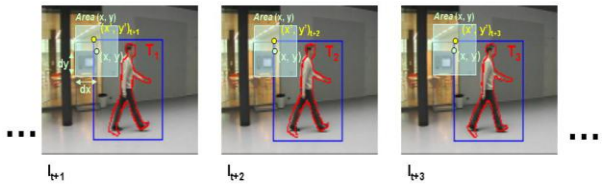


Figure 2:Tracing the Human Gestures

Figure 2 denotes the human body is being tracked by the camera in the 2d postural view. Where S is the mapped image abstract formed image containing n points, and C is the set of contour points in the input image after canny edge detection. Simply relying on the distance between edges produces a lot of false positives, especially in presence of clutter. We therefore also take into account the edge orientation by introducing a penalty term where are the edge orientation respectively at the point si and at the contour point cj , and K is a weight that defines the slope of the penalty function.

Since it follows HAMILTON JACOBIAN formulae for generating codes.

The Hamilton–Jacobin equation is a first-order, non-linear partial differential equation

$$H + \frac{\partial S}{\partial t} = 0$$

where

$$H = H\left(q_1, \cdots, q_N; \frac{\partial S}{\partial q_1}, \cdots, \frac{\partial S}{\partial q_N}; t\right)$$

Figure 3 examines the abstract layer formation for the human which is recognized by the biometric 3d camera, The algorithm used is thinning layer mapping algorithm using point’s generation as discussed above, our template database contains different scale templates.

To allow effective comparison between the chamfer distances for such templates, we explicitly introduce a scale factor k into Equation 1 to normalize the distance to the value that would be computed if the template has not been scaled. Finally we introduce the Tukey robust estimator to reduce the effect of outliers or missing edges. We therefore take Chamfer distance to be.

VI. Implementation Details

In practice, a naive implementation of this method would be computationally very less when compared to other biometrical techniques. Therefore we propose an way of finding the best matches .since the person we are going to refer of recognition will be mapped in various angles form 0 to 360 degrees and database initially .then the mapped image is recognized by framing an abstract for the image. The unique can be proving between a twin pairs. We now prove that the solution of the 3D Heat Problem

$$\begin{aligned} u_t &= \nabla^2 u, \qquad \mathbf{x} \in D \\ u(\mathbf{x},t) &= 0, \qquad \mathbf{x} \in \partial D \\ u(\mathbf{x},0) &= f(\mathbf{x}), \qquad \mathbf{x} \in D \end{aligned}$$

s. Define $v = u_1 - u_2$.

$$V(t) = \int\int\int_D v^2 dV \geq 0$$

$V(t)$ 0 since the integrand v^2 $(\mathbf{x}, t)$ 0 for all $(\mathbf{x}, t)$.

$$\begin{aligned} v_t &= \nabla^2 v, \qquad \mathbf{x} \in D \\ v(\mathbf{x},t) &= 0, \qquad \mathbf{x} \in \partial D \\ v(\mathbf{x},0) &= 0, \qquad \mathbf{x} \in D \end{aligned}$$

Differentiating in time gives

$$\frac{dV}{dt}(t) = \int\int\int_D 2vv_t dV$$

Substituting for v_t from the PDE yields

$$\frac{dV}{dt}(t) = \int\int\int_D 2v\nabla^2 v dV$$

By result,

$$\frac{dV}{dt}(t) = 2\int\int_{\partial D} v\nabla v \cdot \hat{\mathbf{n}} dS - 2\int\int\int_D |\nabla v|^2 dV$$

But on ∂D, $v = 0$, so that the first integral on the r.h.s. vanishes. Thus

$$\frac{dV}{dt}(t) = -2\int\int\int_D |\nabla v|^2 dV \leq 0$$

Also, at $t = 0$,

$$V(0) = \int\int\int_D v^2(\mathbf{x},0)\, dV = 0$$

Thus $V(0) = 0$, $V(t)$ 0 and dV/dt 0, i.e. $V(t)$ is a non-negative, non-increasing function that starts at zero. Thus $V(t)$ must be zero for all time t, so that $v(\mathbf{x}, t)$ must be identically zero throughout the volume D for all time, implying the two solutions are the same, $u_1 = u_2$. Thus the solution to the 3D heat problem is unique. Since the abstract formation is based on RGB value. Reducing down the value of BLUE and GREEN and setting the value of RED as 1 and B,G as 0.here the color values are evaluated based on ultimate value which are present in the composite image of a human body

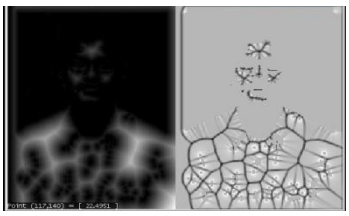


Figure 4: Generating Points to Human Body-Skeletonized View

Figure 4 denotes the point generation for the human image, here the abstract formed image is supposed to generate code by fixing the angles of the image .based on direction since 3D image plays an vital .since an axis reside to rest .2D is taken and calculated by generating points for X,Y axis. Since the code is generated for the points by means of ASCII conversion .then the image is based. The procedure is followed from initial stage to final stage. if the based code and newly generated code matches. Access permitted and body is recognized.

Algorithm: HBR Algorithm

Input: An image A and compared image C.

Output: A Recognized Image R.

Steps:

1. Predict the image and make necessary selection
2. Scan the image and perform abstraction, form abstract for the image with reduced green and blue value to zero and red as one.
3. Make possibilities of finding 3D axis over the images and generate points based on missing axis and plot the values to form the human body skeleton.
4. Generate code for the points and its value.
5. Make comparison of values real time value with data based value
6. Perform authentication & recognition based on code generation

---

VII. Results

We have already shown some of the results obtained from several image sequences with cluttered background. Note that the subjects move closer or further so that their apparent scale changes and turn so that the angle from which they are seen also varies. All the templates in our database are rendered from virtual cameras that are positioned at 1.20m from the ground level, so that optimal results can be expected when the camera is at that height. However, our algorithm is very robust with respect to camera position. Shows its good behavior even when the kinect camera is placed high above the head of the person. In further demonstrate that the detections are correct even when the edge images are very noisy. Furthermore, we can map the detected templates back to full 3D poses as shown in. Remember that our method is designed to detect people in a specific pose. As shown in the walking sequences of, that is

exactly what it does. Note that the camera moves to follow the person. Typical failure modes involve a detection location that is usually correct but an inaccurate orientation or scale, as shown in. To quantify this, we estimated the error rate on the two movies supplied with the paper as supplementary

Figure 5: Full Composite Recognized Image of a Human Gesture

Figure5 denotes the full composite image of the recognized human body .In summary, our method detects people in the target posture with a very low error rate. The few false positives still correspond to people but at somewhat inaccurate scales or orientations. While this paper focuses on pure detection it is therefore clear that the performance of our algorithm could be further increased by points generating.

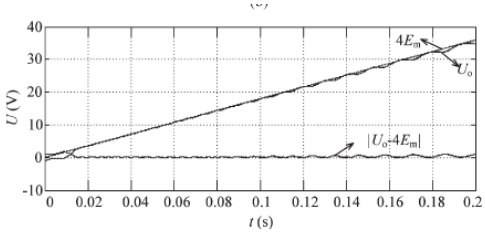


Figure 6:Graph:Corrected Error Rate of Two Movies in Recognizing Human Posture

Here the graph denotes the actual prediction of human body with corrected error rate (U0) in kinect at the rate of 4Em, We have presented a method for human body pose detection that combines points matching and motion information in an original way. This is important because human motion is very different from other kinds of motions and can be effectively used to reduce the false positive and negative detection rate. As a result, we have

VIII. CONCLUSION

We have presented a method for human body pose detection that combines points matching and motion information in an original way. This is important because human motion is very different from other kinds of motions and can be effectively used to reduce the false positive and negative detection rate. As a result, we have already been able to demonstrate very good results for indoor and outdoor sequences for which, Background subtraction is impossible, under difficult lighting conditions, different camera viewpoints and apparent scale changes. Furthermore, since the detected templates are projections of 3-D models, mapping them back from 2-D to full 3-D poses is straightforward. Our approach, even though tested on specific human motion, is generic and could be applied for any other actions that all people perform in roughly similar ways but with Substantial individual variations. The only requirement is that a representative motion database can be built.

This method, with its accurate 3D pose detections, is a key step towards robust full 3-D body pose tracking algorithms that can initialize and re-initialize themselves in difficult real-world conditions where techniques such as background.

REFERENCES

[1] A. Agarwal and B. Triggs. 3d human pose from silhouettes by relevance vector regression. In Conference on Computer Vision and Pattern Recognition, 2008.

[2] A.A. Efros, A.C. Berg, G. Mori, and J. Malik. Recognizing action at a distance. In International Conference on Computer Vision, Pp 726–733, October 2010.

[3] A. Elgammal and C.S. Lee. Inferring 3D Body Pose from Silhouettes using Activity Manifold Learning. In CVPR, Washington, DC, June 2008.

[4] R. Fablet and M.J. Black. Automatic Detection and Tracking of Human Motion with a View-Based Representation. In European Conference on Computer Vision, May 2007.

[5] P. Felzenszwalb and D. Huttenlocher. Pictorial Structures for Object Recognition. International Journal of Computer Vision, 16(1), 2005.

[6] Chen Wu and Aghajan, H. Model-based human posture estimation for gesture analysis in an opportunistic fusion smart camera network, 2007. AVSS 2007.

[7] Y. Azoz, L. Devi, and R. Sharma, "Vision-Based Human Arm Tracking for Gesture Analysis Using Multimodal Constraint Fusion,"Proc. 1997 Advanced Display Federated Laboratory Symp., Adelphi, Md., Jan. 1997.

[8] R. Bajcsy, "Active Perception," Proc. IEEE, Vol. 78, Pp. 996-1,005, 1988.

[9] T. Baudel and M. Baudouin-Lafon, "Charade: Remote Control of Objects Using Free-Hand Gestures," Comm. ACM, Vol. 36, No. 7, Pp. 28-35, 1993.

[10] D.A. Becker and A. Pentland, "Using a Virtual Environment to Teach Cancer Patients T'ai Chi, Relaxation, and Self-Imagery," Proc. Int'l Conf. Automatic Face and Gesture Recognition, Killington,Vt. , Oct. 1996.

[11] A. Blake and A. Yuille, Active Vision. Cambridge, Mass.: MITPress, 1992.

[12] Lee.W. Campbell, Aaron F. Bobick "Recognition of Human body motion using phase space constraints"-International conference on computer vision-ICCV, Pp.624-630.,1995

[13] H.A. Boulard and N. Morgan, Connectionnist Speech Recognition. AHybrid Approach. Norwell, Mass.: Kluwer Academic Publishers, 1994.

[14] U. Bröckl-Fox, "Real-Time 3D Interaction With Up to 16 Degrees of Freedom From Monocular Image Flows," Proc. Int'l Workshop on Automatic Face and Gesture Recognition, Zurich, Switzerland, Pp. 172-178, June 1995.

[15] L.W. Campbell, D.A. Becker, A. Azarbayejani, A.F. Bobick, and A. Pentland, "Invariant Features for 3D Gesture Recognition, "Proc. Int'l Conf. Automatic Face and Gesture Recognition, Killington, Vt., Pp. 157-162, Oct. 1996.

[16] C. Cedras and M. Shah, "Motion-Based Recognition: A Survey, "Image and Vision Computing, Vol. 11, Pp. 129-155, 1995.

[17] K. Cho and S.M. Dunn, "Learning Shape Classes," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 16, Pp. 882-888, Sept.1994.

[18] R. Cipolla and N.J. Hollinghurst, "Human-Robot Interface byPointing With Uncalibrated Stereo Vision," Image and Vision Computing, Vol. 14, Pp. 171-178, Mar. 1996.

[19] R. Cipolla, Y. Okamoto, and Y. Kuno, "Robust Structure From Motion Using Motion Parallax," Proc. IEEE Int'l Conf. Computer Vision, Pp. 374-382, 1993.

[20] E. Clergue, M. Goldberg, N. Madrane, and B. Merialdo, "Automatic Face and Gestural Recognition for Video Indexing," Proc. Int'l Workshop on Automatic Face and Gesture Recognition, Zurich, Switzerland, Pp. 110-115, June 1995.

[21] T. F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham, "Active Shape Models—Their Training and Application," Computer Vision and Image Understanding, Vol. 61, Pp. 38-59, Jan. 1995.
[22] J.L. Crowley, F. Berard, and J. Coutaz, "Finger Tacking As an Input Device for Augmented Reality," Proc. Int'l Workshop on Automatic Face and Gesture Recognition, Zurich, Switzerland, Pp. 195-200, June 1995.
[23] Y. Cui and J. Weng, "Learning-Based Hand Sign Recognition," Proc. Int'l Workshop on Automatic Face and Gesture Recognition, Zurich, Switzerland, pp. 201-206, June 1995.
[24] Y. Cui and J.J. Weng, "Hand Segmentation Using Learning-Based Prediction and Verification for Hand Sign Recognition," Proc. Int'lConf. Automatic Face and Gesture Recognition, Killington, Vt., Pp. 88-93, Oct. 1996.
[25] Cuong Tran , Anup Doshi, Mohan Manubhai Trivedi "Modeling and prediction of driver behavior by foot gesture analysis "Proc. Int'l Workshop on Laboratory for Intelligent and Safe Automobiles (LISA), University of California, San Diego, CA 92093, USA
[26] A.F. Bobick and J.W. Davis, "Real-Time Recognition of Activity Using Temporal Templates," Proc. Int'l Conf. Automatic Face and Gesture Recognition, Killington, Vt., Oct. 1996.