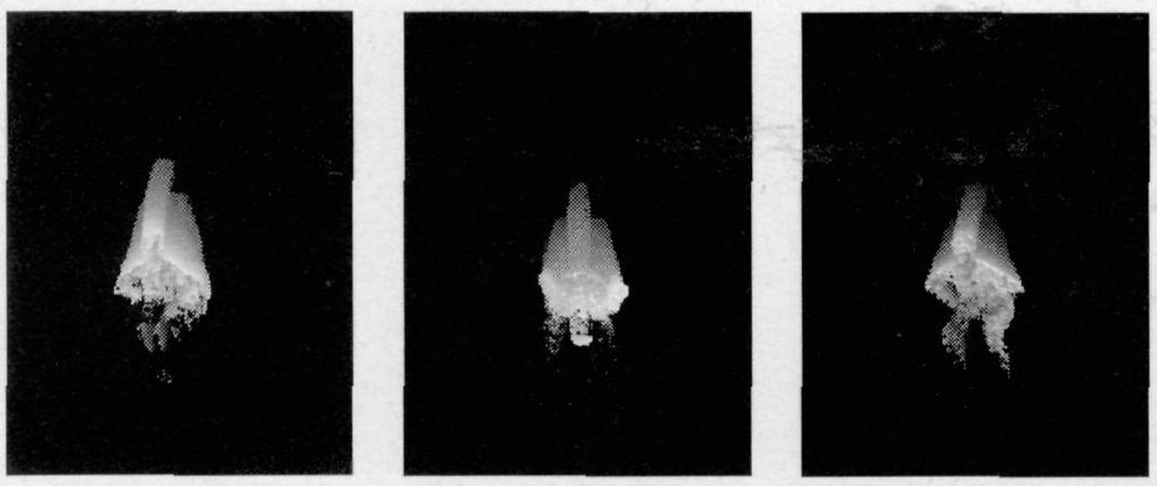
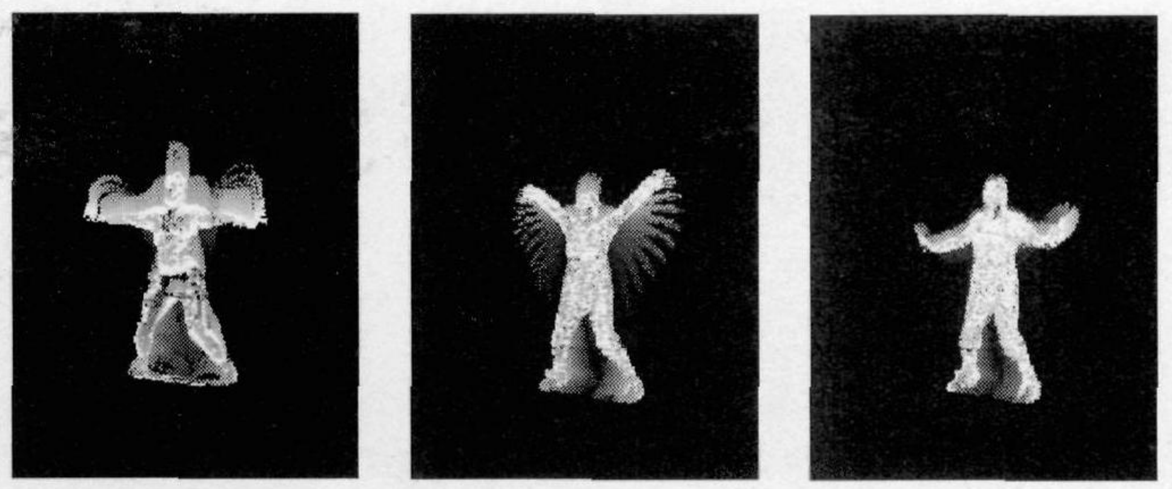


(a) (b) (c)

Fig. 7. An example of MHIs with similar statistics. (a) Test input of move 13 at 30°. (b) Closest match which is move 6 at 0°. (c) Correct match.



(a) (b) (c)

Fig. 8. Example of error where failure is caused, by both the inadequacy of using image differencing to estimate image motion and the lack of the variance data in the recognition procedure. (a) Test input of move 16. (b) Closest match which is move 15. (c) Correct match.

template with similar statistics. For example, consider sitting and crouching motions when viewed from the front. The observed motions are almost identical and the coarse temporal template statistics do not distinguish them well.

4.3 Combining Multiple Views

A simple mechanism to increase the power of the method is to use more than one camera. Several approaches are possible. For this experiment, we use two cameras placed such that they have orthogonal views of the subject. The recognition system now finds the minimum sum of Mahalanobis distances between the two input templates and two stored views of a movement that have the correct angular difference between them, in this case 90°. The assumption embodied in this approach is that we know the approximate angular relationship between the cameras.

TABLE 2
Results Using Two Cameras Where the Angular Interval Is Known and Any Matching Views Must Have the Same Angular Distance

		Closest Dist	Closest Move	Correct Dist	Median Dist	Rank
Test	1	2.13	1	2.13	6.51	1
	2	12.92	2	12.92	19.58	1
	3	7.17	3	7.17	18.92	1
	4	1.07	4	1.07	7.91	1
	5	16.42	5	16.42	32.73	1
	6	0.88	6	0.88	3.25	1
Test	7	3.02	7	3.02	7.81	1
	8	36.76	8	36.76	49.89	1
	9	5.10	8	6.74	8.93	3
	10	0.68	10	0.68	3.19	1
	11	1.20	11	1.20	3.68	1
	12	2.77	12	2.77	15.12	1
Test	13	0.57	13	0.57	2.17	1
	14	6.07	14	6.07	16.86	1
	15	2.28	15	2.28	8.69	1
	16	1.86	15	2.35	6.72	2
	17	2.67	8	3.24	7.10	3
	18	1.18	18	1.18	4.39	1

Table 2 provides the same statistics as the first table, but now using two cameras. Notice that the classification now contains only three errors. The improvement of the result reflects the fact that for most pairs of this suite of movements, there is some view in which they look distinct. Because we have 90° between the two input views, the system can usually correctly identify most movements.

We mention that if the approximate calibration between cameras is not known (and is not to be estimated) one can still logically combine the information by requiring consistency in labeling. That is, we remove the interangle constraint, but do require that both views select the same movement. The algorithm would be to select the move whose Mahalanobis sum is least, regardless of the angle between the target views. If available, angular order information—e.g., camera 1 is to the left of camera 2—can be included. When this approach is applied to the aerobics data shown here, we still get similar discrimination. This is not surprising because the input views are so distinct.

To analyze the remaining errors, consider Fig. 8, which shows the input for move 16. Left to right are the 30° MHIs for the input, the best match (move 15), and the correct match. The test subject performed the move much less precisely than the original aerobics instructor. Because we were not using a Mahalanobis variance across subjects, the current experiment could not accommodate such variation. In addition, the test subject moved her body slowly while wearing low frequency clothing resulting in an MHI that has large gaps in the body region. We attribute this type of failure to our simple (i.e., naive) motion analysis; a more robust motion detection mechanism would reduce the number of such situations.

5 Segmentation and Recognition

The final element of performing recognition is the temporal segmentation and matching. During the training phase, we measure the minimum and maximum duration that a movement may take, τ_{min} and τ_{max}. If the test motions are performed at varying speeds, we need to choose the right τ for the computation of the MEI and the MHI. Our current system uses a backward looking variable time window. Because of the simple nature of the replacement operator, we can construct a highly efficient algorithm for approximating a search over a wide range of τ.