HUMAN MOVEMENT UNDERSTANDING: A VARIETY OF PERSPECTIVES

Norman I. Badler
Joseph O'Rourke
Stephen Platt
Mary Ann Morris

Department of Computer and Information Science
Moore School D2
University of Pennsylvania
Philadelphia, PA 19104

ABSTRACT

Our laboratory is examining human movement from a variety of perspectives: synthesis of animated movements and analysis of moving images. Both gestural (limb and hand) and facial movements are being explored.

Key Words and Phrases: Human movement, movement representation, motion analysis, computer vision, facial expression, facial analysis, constraint network.

I HUMAN MOVEMENT SYNTHESIS

Our laboratory is examining human movement representation from a variety of perspectives, including synthesis of three-dimensional animated movements and analysis of moving images. These broad areas are further refined into gestural (limb and hand) and facial movements since very different sorts of actions take place in jointed skeletal movement and "rubber-sheet" facial distortions. Our human body model [5](Fig. 1) and hand model [2](Fig. 2) are based on spherical decompositions of three-dimensional objects. Our goals have been to develop representations for human movements which permit both analysis and synthesis, are "complete" in terms of the range of human movements, and yet provide a "high-level" interface suitable for specifying or describing movement. Many of these issues are addressed in a recent survey by Badler and Smoliar [6], so we shall emphasize only the more recent work in human motion analysis and facial expression synthesis.

II HUMAN MOVEMENT ANALYSIS

There have been a rather small number of attempts to analyze complex movement presented in time-varying images. Rashid [13] uses moving light displays to track body joints and infer three-dimensional connections. Tsotsos [14] describes non-articulated shape changes. Badler [1] attempts conceptual descriptions of rigid jointed movements. Recently O'Rourke [9,10] describes a computer system which accepts a sequence of two-dimensional images of a solid, three-dimensional body [5] performing some motion sequence (Fig. 3). The output of the system is a description of the motion as coordinate-versus-time trackings of all body joints and as movement instructions suitable for controlling the simulation of a human body model [3,4,15]. The simulation includes a detailed model of a human body which incorporates the structural relationships between the body parts and the physical limitations to relative movement between the parts. This model is used to guide the image analysis through a prediction/verification control cycle. Predictions are made at a high level on the basis of previous analysis and the properties of the human model. The low level image analysis then verifies the prediction and the model is adjusted according to any discrepancies found or new knowledge acquired. This cycle is repeated for each input frame.

The information extracted from the image is integrated into the current position of the model through a constraint network operating on real-valued coordinate regions in three-dimensional space. Possible positions of body features are represented as unions of orthogonally-oriented rectangular boxes. Relationships among body parts-for example, distance constraints imposted by the body skeleton-are enforced in the network. As new joint positional information is extracted from the image it is added to the network and its geometrical consequences immediately propagated throughout the network. Only head, hands, and feet are located in the image space, yet all remaining body joints may be tracked by the geometric inference process in the network. Figure 4 shows the constraint boxes for each joint of the body given the moving images of Fig. 3.

III FACIAL ANIMATION

We are also investigating the representation and simulation of the actions performable on the human face. The face presents an interesting problem for simulation, as it is composed of mostly independent sections of non-rigid masses of skin and muscle, such as the areas around the eyes, mouth, and cheeks. This type of action is basically different from gross body movement in that a facial action will affect the visible results of other actions performed in the same area of the face.

Our internal representation of the face is based on FACS, the Facial Action Coding System [7]. The system categorizes basic actions performable and recognizable (by a human notator) on the face. It is also easily translated into a state-description of the face, in terms of muscle contractions. A complete recognition and simulation system for the face would consist of a camera, computer processing to obtain an idealized internal representation of the action, and a simulation of the action performed on a graphic output device. Once the camera image is obtained, analysis is performed to produce the AU (FACS Action Unit) state of the face. This analysis is relatively simple, as it consists of identifying the presence/absence of "features" such as wrinkles and bulges on the face. (Note that this analysis does not require "recognition" of a particular face, just good comparison between successive images of the same face.) The current technique under investigation uses an improved method of "rubber-sheet" matching [8]. Each AU effects only a small set of muscles; their union gives the muscle-status of the face. The specified muscle contractions are then simulated on the face.

The face is represented by a network of points and interconnecting arcs (Figs. 5 and 6) [12]. It also has a higher level of organization which partitions the network into the skin surface and specific muscles. (It is this muscle organization which distinguishes our work from that of Parke [11].) The skin is a "warped" two-dimensional surface of points and arcs. The points represent the basic surface, while the arcs contain information specific to their locale, such as the elasticity ("stretchiness") of the skin between the arc's points. Stretching the skin (by contracting a muscle) causes first local motion, followed by propagation of the skin distortion. Muscles are also points, connected to but beneath the skin. They are distinguished by being the initiation of the distortion of the skin surface. An AU is thus merely a set of muscles, with appropriate magnitudes of initial force of contraction.

IV FUTURES

Our research into human movement understanding has the joint goals of achieving conceptual descriptions of human activities and producing effective animations of three-dimensional human movement from formal specification or notation systems [6]. One application of this effort is in the synthesis and analysis of American Sign Language. A prototype synthesizer for ASL is being designed to facilitate experimentation with human dynamics and movement variations which have linguistic import for facial and manual communication.

ACKNOWLEDGEMENTS

The support of NSF Grants MCS76-19464, MCS78-07466, and O'Rourke's IBM Fellowship are gratefully acknowledged.

REFERENCES

[1] Badler, N.I. Temporal scene analysis: Conceptual descriptions of object movements. Tech. Rep. 80, Dept. of Computer Science, Univ. of Toronto, Feb. 1975.

[2] Badler, N.I. and J. O'Rourke. Representation of articulable, quasi-rigid, three-dimensional objects. Proc. Workshop on Representation of Three-Dimensional Objects, Philadelphia, PA, May 1979.

[3] Badler, N.I., J. O'Rourke, and B. Kaufman. Special problems in human movement simulation, Computer Graphics Summer 1980.

[4] Badler, N.I., J. O'Rourke, S.W. Smoliar, and L. Weber. The simulation of human movement by computer. Tech. Rep. Dept. of Computer and Information Science, Univ. of Pennsylvania, July 1978.

[5] Badler, N.I., J. O'Rourke, and H. Toltzis. A spherical representation of a human body for visualizing movement. IEEE Proceedings 67: 10 (1979) pp. 1397-1403.

[6] Badler, N.I. and Smoliar, S.W. Digital representations of human movement. Computing Surveys 11: 1 (1979) pp. 19-38.

[7] Ekman, P. and W. Friesen. Facial Action Coding System. Consulting Psychologists Press, Palo Alto, CA. 1978.

[8] Fischler, M.A. and R.A. Erschlager. The representation and matching of pictoral structures. IEEE Tr. on Computers C-22: 1 (1973).

[9] O'Rourke, J. Image analysis of human motion. Ph.D. Diss. Dept. of Computer and Information Science, Univ. of Pennsylvania, 1980.

[10] O'Rourke, J. and N.I. Badler. Human motion analysis using constraint propagation. To appear IEEE-PAMI November 1980.

[11] Parke, F.I. Animation of faces. Proc. ACM Annual Conf. 1972, pp. 451-457.

[12] Platt, S. Animating facial expressions. MSE Thesis, Dept. of Computer and Information Science, Univ. of Pennsylvania, 1980.

[13] Rashid, R. Lights: A system for the interpretation of moving light displays. To appear IEEE-PAMI, Nov. 1980.

[14] Tsotsos, J. A framework for visual motion analysis. Ph.D. Diss. Dept. of Computer Science, Univ. of Toronto, 1980.

[15] Weber, L., S.W. Smoliar, and N.I. Badler. An architecture for the simulation of human movement. Proc. ACM Annual Conf. 1978, pp. 737-745.

Figure 1. 3D human body model.

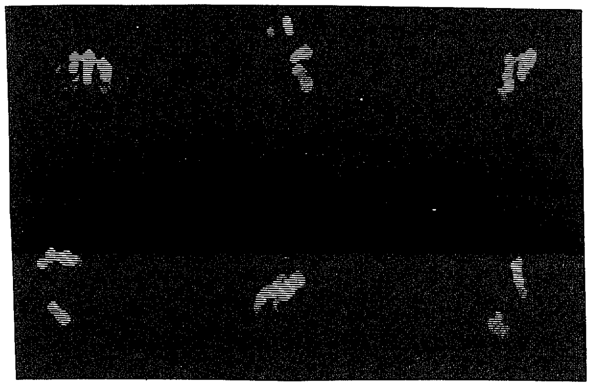

Figure 2. 3D hands in sign language positions.

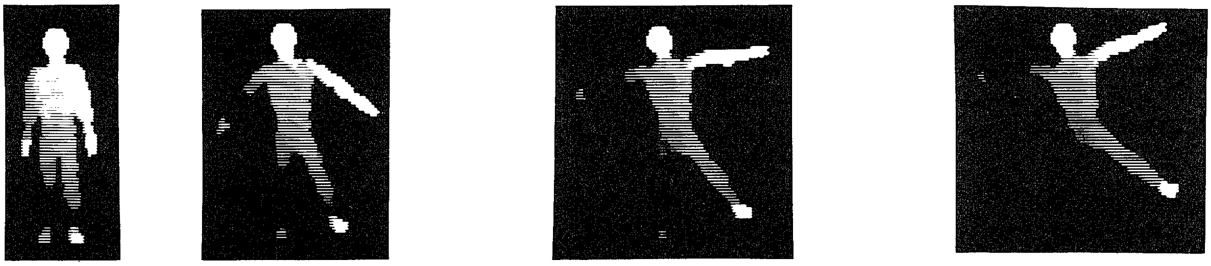

Figure 3. Simulated movement for analysis.

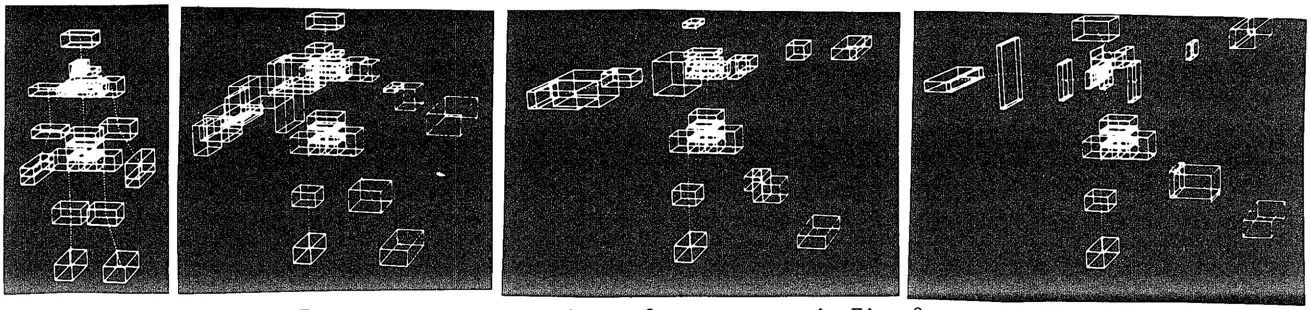

Figure 4. Constraint boxes for movements in Fig. 3.

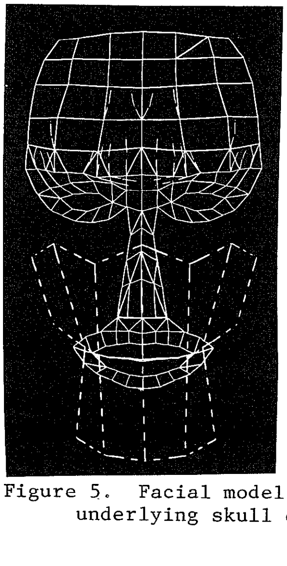

Figure 5. Facial model showing some of the underlying skull outline and muscles.

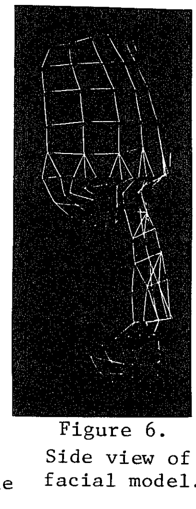

Figure 6. Side view of facial model.

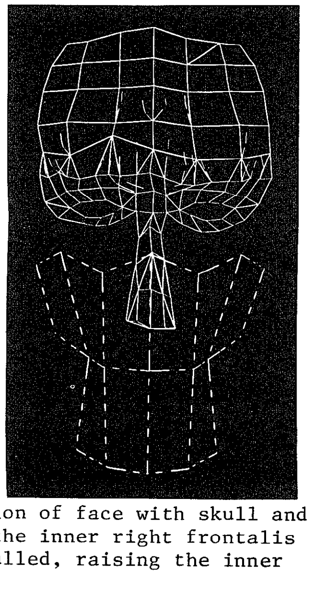

Figure 7. Upper portion of face with skull and muscles; the inner right frontalis has been pulled, raising the inner right brow.