# FLOW LOOKUP AND BIOLOGICAL MOTION PERCEPTION

*Nicholas R. Howe*

Smith College
Northampton, Massachusetts, USA

## ABSTRACT

Optical flow in monocular video can serve as a key for recognizing and tracking the three-dimensional pose of human subjects. In comparison with prior work using silhouettes as a key for pose lookup, flow data contains richer information and in experiments can successfully track more difficult sequences. Furthermore, flow recognition is powerful enough to model human abilities in perceiving biological motion from sparse input. The experiments described herein show that a tracker using flow moment lookup can reconstruct a common biological motion (walking) from images containing only point light sources attached to the joints of the moving subject.

## 1. INTRODUCTION

Human beings rely on vision to determine where other people are and what they are doing, both in the real world and in media such as television and film. By comparison, computers fall short even in simply tracking the three-dimensional pose of a human being moving in a short video. Although current articulated pose trackers can follow very simple movements, they make errors when presented with more difficult sequences [2]. They also may require more coddling: some systems rely on multiple cameras, human initialization, or *a priori* appearance models, and they typically require large amounts of computational time. Clearly, great room exists for improvement.

Intense research interest has focused to date on trackingof human pose, as evidenced by a recent survey [7]. Some of the newest approaches, based upon lookup or recognition of known poses from their silhouettes, have shown great success in tracking both hands [9, 13] and entire humans [4, 8]. However, silhouettes suffer from several drawbacks when used as the sole source of pose information. Most notably, they cannot provide feedback about body parts that do not abut the silhouette border. More subtly, silhouettes provide limited cues about the movement direction of rotating bodies. To address these shortcomings, this paper investigates optical flow as a cue for recognition-based tracking, either alone or in conjunction with silhouette cues. In contrast to silhouettes, optical flow provides strong information about rotating bodies, and can also reveal the motion of body parts whose silhouettes are entirely surrounded by others. Its strengths therefore complement those of silhouettes. Although optical flow has been used for human detection and verification [11, 14], it has received little attention as a general means of tracking pose.

Using flow to recognize pose (i.e., *flow lookup*) proves interesting in another context as well. Psychological evidence dating back to the 1970s indicates that people can recover the pose of familiar biological forms in motion [5] from extremely sparse data, specifically images consisting only of isolated points that move as though attached to the joints of an articulated figure. Although this phenomenon has been explained as a bottom-up detection of pairs of rigidly-connected points leading to the construction of a human figure [3], this theory cannot account for details such as orientation-specificity. (Upside-down walkers are far more difficult to interpret than those in a normal orientation [12].) This paper develops the alternative possibility that human observers rely directly on the optical flow of the points in recognizing biological motion. Flow lookup seems to offer a convenient explanation for these phenomena, if it can indeed recover human motion from images of moving point sets. The experiments presented in Section 3 indicate that it can.

The remainder of the paper is organized as follows. Section 2 describes the flow lookup algorithm and its implementation, and Section 3 presents related experiments. Section 4 concludes with a discussion and possible future work.

## 2. ALGORITHM

This work adopts the general approach used for other recognition-based trackers, summarized hereafter for the reader's convenience. The video input undergoes some preprocessing to extract relevant features from each image frame (e.g., silhouettes or flow fields). These features become the keys used to retrieve known poses from a library. Because the library will typically not contain an exact match to the observed pose, and because the extracted features may not clearly differentiate the true pose from other poses with similar feature values, a collection of potential poses may be retrieved for each frame. This guards against situations where the correct pose is not the top-ranked hit using the chosen feature set. Once a small collection of potential poses has been identified for each frame, the collection of observations forms a temporal Markov chain, and the Viterbi algorithm (forward-backward chaining) can find the sequence of poses that minimizes an objective function. Typically, the objective function chosen will have both "smoothness" and "data" terms, and penalizes sequences that change pose sharply between adjacent frames or do not closely match the observations. This paper adopts an objective function used in previous work [4], and follows that work in other details except as noted below.

### 2.1. Working with flow

Three issues must be addressed before optical flow can be used for recognition. First, the flow must be measured in the input video. Second, flow information must also be associated with each library pose. Finally, a method must be chosen for comparing the observed flows to the library flows.

Flow measurement techniques have received extensive research attention that need not be replicated here, since comprehensive surveys are available [1]. Although many methods could serve equally well, this work uses Krause's algorithm based upon polynomial fitting [6], which produces a 2-d flow vector for each pixel. Under the assumption of a static camera, flow in background regions is set to zero to suppress noise.

Adding flow information to the pose library is not difficult, assuming that the library is populated using data gathered in a motion capture studio. The captured motion can generate reasonable flow values if a model of body shape is available (as would be needed to generate artificial silhouettes). To infer the flow, simply compute the position of corresponding points visible on the body surface in two successive frames, and then determine the resultant observed motions in the camera image plane. This method has been used elsewhere [14] and produces flows that look quite realistic.

When selecting a measure for flow comparison, storage and processing requirements must be considered. Retaining every flow field in its entirety would require large amounts of storage and extensive processing during lookup. Instead, this work compares flow fields based upon standardized central moments of each component of the flow field (see Figure 1). This simultaneously reduces storage requirements and increases retrieval speed. (Note that the flow moments used need not be rotationally invariant because gravity imposes a fixed orientation on the tracked scene. A human figure upside-down is clearly not equivalent to one rightside-up.) Euclidean distance serves for the comparison of the moment vectors from two flow fields.

$$m_{00} = \frac{\sum F_x(x,y)}{|P|} \tag{1}$$

$$m_{10} = \frac{\sum (x-\overline{x}) F_x(x,y)}{\sum |x-\overline{x}|} \tag{2}$$

$$m_{01} = \frac{\sum (y-\overline{y}) F_x(x,y)}{\sum |y-\overline{y}|} \tag{3}$$

$$m_{20} = \frac{\sum (x-\overline{x})^2 F_x(x,y)}{\sum |x-\overline{x}|^2} \tag{4}$$

$$m_{11} = \frac{\sum (x-\overline{x})(y-\overline{y}) F_x(x,y)}{\sum |(x-\overline{x})(y-\overline{y})|} \tag{5}$$

$$m_{02} = \frac{\sum (y-\overline{y})^2 F_x(x,y)}{\sum |y-\overline{y}|^2} \tag{6}$$

**Fig. 1**. Equations used to compute moments of the $x$ component of a flow $F$. All sums are over $P$, the set of coordinates $(x, y)$ of foreground pixels. A similar set of equations yields six moments of the $y$ component $F_y$.

### 2.2. Flow fields from point sources

Although standard practice guides the computation of optical flow from ordinary video, the moving points of light used in research on *biological motion displays* require special treatment. To serve as a model of biological vision, an algorithm can use no information or assumptions about the input not available to the human visual system. In particular, the observed point locations must *not* be used directly to infer parameters of an articulated human body model. In the experiments to follow, the moving point images serve solely to infer a series of optical flow fields.

Flow information is extrapolated from points where it is known (the visible moving points) to regions where it is unknown (the dark background) via a simple heuristic. The flow at any given location is taken as the mean flow of the nearby visible points, weighted by a Gaussian function of their distance. Furthermore, flow at locations farther than a threshold distance from any visible point is set to zero. Providing that flow discontinuities are rare, and that the visible points are well distributed over the subject, this heuristic approach produces an adequate estimate of the flow field.

## 3. EXPERIMENTS

The experiments that appear below assess the promise and performance of flow lookup for three-dimensional pose tracking. As a validation of the technique, the first experiment shows that flow lookup alone can track a circular walking example as well as or better than silhouette lookup. A second experiment combines flow and silhouette lookup. The final set of experiments focuses on biological motion displays. Under the assumptions detailed above regarding

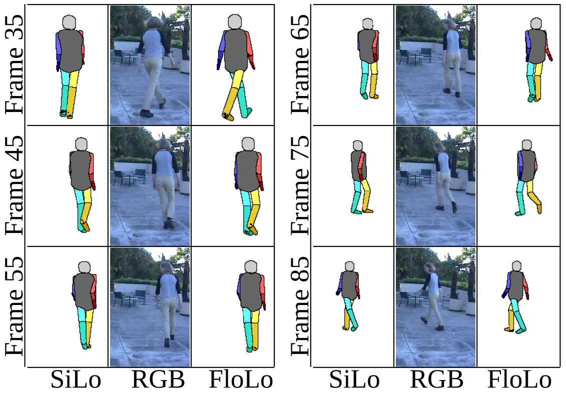

**Fig. 2**. Improved tracking of *CircleWalk* clip (closeup). The silhouette-based tracker loses the legs in frames 50-80, but the flow-based tracker does not.

the conversion of moving point data into flow fields, the results show that flow lookup can successfully recover the underlying motion from this sparse data source, when the motion is familiar (i.e., similar to a set of known examples).

### 3.1. Flow lookup for standard video

Other work has examined the strengths and weaknesses of trackers based upon silhouette lookup [4]. This section therefore compares such *SiLo* trackers with those based upon flow lookup (*FloLo*). An existing SiLo implementation [4] serves as the control.

The *CircleWalk* clip forms a good basis for comparison, because it includes walking motions from changing points of view, and has appeared in several previous works [4, 10]. It shows a woman walking in a circular pattern, for about 140 frames. Figure 2 displays the results in a crucial segment. (Video animations of this and subsequent clips may be found at the author's web page, `http://www.cs.smith.edu/˜nhowe/research/flowtrack`.) The SiLo tracker loses track of the legs as the subject's direction of travel becomes aligned with the line-of-sight axis (around frame 50) and does not regain it again until around frame 80. This is partly due to a failure to separate the legs during silhouette extraction, but mostly may be attributed to fact that changes to the silhouette boundary are small and nondescript during this period. The small frame-to-frame changes allow the smoothness term to dominate the Markov chain solution, and the pose appears frozen in place. By comparison, the FloLo tracking result has no trouble with this portion of the sequence, and matches every movement of the legs. On the other parts of the clip, the two methods give roughly comparable (good) results.

For complex motions, combining flow and silhouette lookup yields better tracking of rotational motions. (This hybrid method works by pooling the sets of candidates generated by each lookup technique.) The *Dancer* [4] clip contains a difficult 180 degree leaping turn. Although the SiLo tracker alone can capture the gist of this motion (using a library of dance poses), it gets the direction of rotation backwards on the turn. Adding flow lookup corrects the rotation direction, as shown in Figure 3.

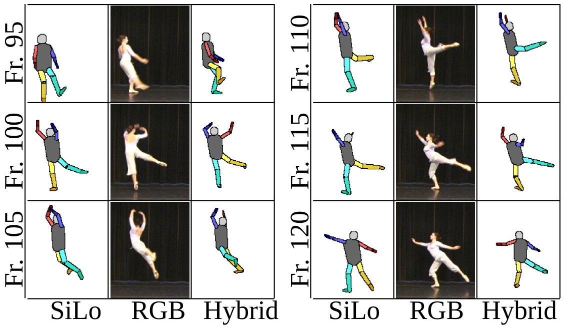

**Fig. 3**. Improved tracking of *Dancer* clip (closeup). The hybrid tracker correctly identifies the direction of rotation.

Some caution is necessary when using flow lookup. As is the case with silhouette extraction, the algorithms that calculate optical flow can make errors. This results in a set of candidate poses retrieved from the library using a flawed query. If the error is small, the set returned by the flawed query will be close enough to the correct pose that smoothing techniques can correct for it. If the error is large, then it may take a number of frames to recover (perhaps ten or so [4]). Second, the size of the pose library needed for flow lookup may be larger than for silhouette lookup, since it is possible for identical silhouettes to have different flows. The library must include representatives of all the motions to be recognized in order for the lookup approach to succeed.

### 3.2. Flow lookup for biological motion displays

The biological motion display used in this experiment comes from previously unseen motion capture data. A single walking sequence of 300 frames provides the spatial locations of selected body points (fifteen joints plus five limb termini) in each frame. Rendering these points from a particular viewpoint provides a synthetic biological motion display. Figure 4 shows results for the classic side-view walker. (Similarly successful results using a rear view of a walker are not shown due to space constraints.) Although the clip contains several walking cycles, the results are nearly identical and only one cycle is shown. The flow lookup algorithm correctly reproduces all the fundamental features of the motion. The tracked figure walks in step with the original, although the reconstruction is a little jerkier, with the

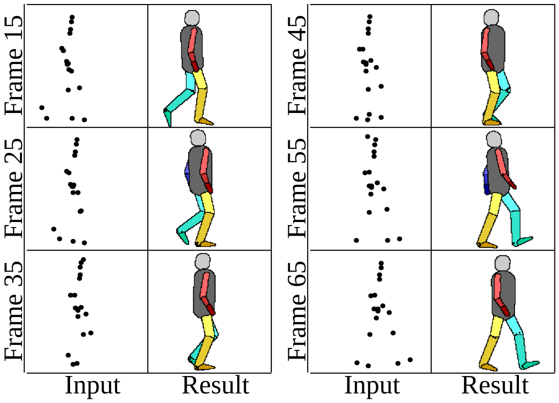

**Fig. 4**. Reconstruction of biological motion from point displays: Sideways walking.

feet tending to stay together longer. This may be due in part to the mechanism chosen for generating the regional flow field: when points with different velocities are close together (as when the feet cross), they tend to average out and cancel each other. Nevertheless, this difficulty did not cause the system to lose track of any limbs, and the reconstructed motion is quite clearly a walk. By contrast, a control experiment (also not shown) using a non-walking point display input generates an unrecognizable result. Thus the algorithm can recognize walking behavior when it is present, but will not hallucinate it when not present.

## 4. CONCLUSION

The results in this paper demonstrate that flow lookup can be a powerful tool for recovering the three-dimensional pose and motion of a familiar articulated figure such as a human being. In particular, the experiments show that this mechanism is powerful enough to duplicate the ability of human beings to recognize familiar motion given extremely sparse point-motion input. Whether humans employ similar techniques or some other mechanism has not been established, but should lead to interesting future research.

As a technique for use in computer analysis of standard video, flow lookup also shows promise. It produces qualitatively better tracking results than silhouette lookup for some motions where silhouette lookup fails. Furthermore, it can combine with silhouette lookup to form a hybrid algorithm employing a pool of potential poses chosen via both techniques. In this manner, each mechanism provides a backup in case the other should fail, and their different strengths can complement each other.

## 5. ACKNOWLEDGEMENTS

*This material is based upon work supported by the National Science Foundation under Grant No. IIS-0328741. The data used in this project were obtained from mocap.cs.cmu.edu. The database was created with funding from NSF EIA-0196217.*

## 6. REFERENCES

[1] J. Barron, D. Fleet, and S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, 1994.

[2] D. DiFranco, T.-J. Cham, and J. M. Rehg. Reconstruction of 3-d figure motion from 2-d correspondences. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 307–314, 2001.

[3] D. Hoffman and B. Flinchbaugh. The interpretation of biological motion. *Biological Cybernetics*, 42:195–204, 1982.

[4] N. Howe. Silhouette lookup for automatic pose tracking. In *IEEE Workshop on Articulated and Nonrigid Motion*, 2004.

[5] G. Johansson. Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, 14:201–211, 1973.

[6] E. Krause. *Motion Estimation for Frame-Rate Conversion*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, June 1987.

[7] T. B. Moeslund and E. Granum. A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding*, 81(3):231–268, March 2001.

[8] G. Mori and J. Malik. Estimating human body configurations using shape context matching. In *European Conference on Computer Vision*, 2002.

[9] R. Rosales, M. Siddiqui, J. Alon, and S. Sclaroff. Estimating 3d body pose using uncalibrated cameras. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001.

[10] H. Sidenbladh. *Probabilistic Tracking and Reconstruction of 3D Human Motion in Monocular Video Sequences*. PhD thesis, Royal Institute of Technology, Stockholm, 2001.

[11] Y. Song, X. Feng, and P. Perona. Towards detection of human motion. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 810–817, 2000.

[12] S. Sumi. Upside-down presentation of the Johansson moving light pattern. *Perception*, 13:283–286, 1984.

[13] C. Tomasi, S. Petrov, and A. Sastry. 3d tracking = classification + interpolation. In *International Conference on Computer Vision*, pages 1441–1448, 2003.

[14] T. Zhao, R. Nevatia, and F. Lv. Segmentation and tracking of humans in complex situations. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 194–201, 2001.