Motion-based Feature Tracking For Articulated Motion Analysis

Hildegard Köhler
Institute for Algorithms and Cognitive Systems
Universität Karlsruhe
Kaiserstr. 12
76131 Karlsruhe, Germany
koehler@ira.uka.de

Annika Wörner
Institute for Algorithms and Cognitive Systems
Universität Karlsruhe
Kaiserstr. 12
76131 Karlsruhe, Germany
woerner@ira.uka.de

ABSTRACT
Feature-based motion perception shows high robustness against environmental influences, large scalability and is highly adaptable to different situations. Furthermore, the motion of features relative to each other usually gives information about the motion properties of the underlying structure; it allows reconstructing the motion and body structure of a moving object without any prior information regarding its appearance. In this article, we show an extension of the Kanade-Lucas-Tomasi feature tracking method appropriate for feature-based articulated motion detection. The motion-based feature tracking restricts the number of features to track to only the moving ones. This allows focusing subsequent data analysis on a subset of feature candidates. We show that by this method, we get a mean feature correctness of 80% with an acceptable stability over more than 50 frames. We compared the overall runtime of the presented method to the general Kanade-Lucas-Tomasi method, showing a performance increase of more than 70%.

Categories and Subject Descriptors
I.2.10 [**Artificial Intelligence**] Vision and Scene Understanding - *Motion - Texture - Video analysis - Intensity, color, photometry, and thresholding.*

General Terms
Experimentation, Algorithms, Performance

Keywords
Motion based feature tracking, articulated body tracking, human motion analysis.

1. INTRODUCTION
One of the main abilities of the human perception system is to determine the form and the underlying body-structure of any moving object. This perception is mostly independent from any environmental influences like a moving background, but also from the visual representation of the object itself, e.g. its size, color or surface appearance. This ability of biological motion perception is, amongst other things, also based on the perception of features, as has been shown in the experiments of Johansson with moving light displays [6].

In everyday life, experimental moving light displays can be represented by points, lines, corners or any distinguishable, unambiguous visual representation. The geometrical clustering and definition of relations among features can be based on spatial relations, but also on the analysis of motion properties. The differentiation between moving and static features as well as their intuitive alignment and clustering allows the reconstruction of complex structures and their recognition even in suboptimal circumstances and with incomplete visual information.

For practical applications, it is important to first differentiate between moving and static feature points in order to focus on moving regions. To preserve the association over a longer time, it is necessary to observe that moving features can vanish from view, e.g. by overlapping or occlusion and new features can appear. It is therefore necessary to define a dynamic model for the continuous integration of new features and for the rejection of invalid features.

In this context we propose a motion-based Kanade-Lucas-Tomasi feature tracker that works exclusively on moving features. This extension of the usual Kanade-Lucas-Tomasi (KLT) feature tracker allows reducing the overall number of features with the effect that less features have to be taken into account e.g. for articulated body tracking. It also handles the initialization of new features, their temporal tracking as well as the rejection of no longer valid features. With the reduced feature set, it is possible to identify rigid body structures just by their feature motion. Because the way they move is defined by the motion of the underlying body structure and their projection in the 2D image, it is possible to estimate the motion of the underlying rigid body. So it becomes possible to estimate the underlying body structure by analyzing the motion properties of feature points.

2. STATE OF THE ART
Automatic detection and tracking of people in different contexts with the goal of motion analysis and recognition is an important subject in many application areas of computer vision, such as human-computer-interaction, entertainment, surveillance, sports analysis and even medical rehabilitation. The growing importance of this field is clear from the increasing number of surveys touching on this subject ([1], [9] and [10]).

Feature-based human motion detection and analysis in this context is mainly based on marker tracking [3], because predefined marker positions usually allow direct reconstruction of the underlying skeleton [12]. This type of motion analysis is mainly used in diagnostic applications for professional sports and medical rehabilitation ([7], [4] and [17]), because here a high data quality is needed, what justifies the effort of using markers.

Feature-based human motion detection without predefined markers is described in [14], [15]. Most work in this area is based on experiments of moving light displays from Johansson [6]. But the important role of feature properties for motion recognition is still not finally defined and also an active research area in neuroscience [5].

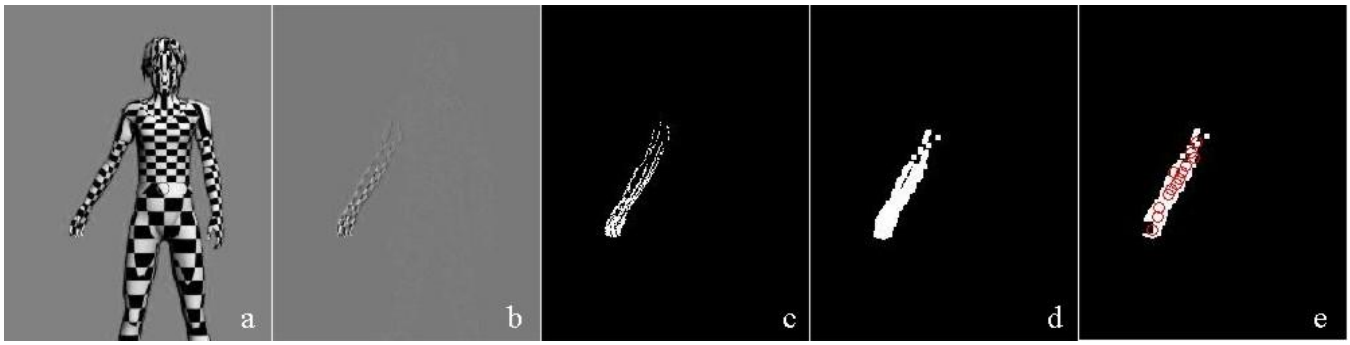

Figure 1: Levels of motion-based feature tracking: a) Original image I(i+1), b) Difference image of I(i) and I(i+1), c) Binarized difference image, d) Dilated difference image, e) Final mask image with new feature points

3. THEORETICAL APPROACH

3.1 Motion-based Feature Tracking

The idea behind motion-based feature tracking is that all features which do not underlie moving regions are not of interest. Furthermore, it is easy to see that a moving region can be detected only by a change of color, brightness or intensity in its region. To put it in another way, in a region with constant color nothing changes, so nothing moves. To explain this phenomenon, one can imagine a monochrome object moving across a contrastive background. The only way to perceive this motion is through the change of color around the borders of the object. This assumption holds a fortiori for intensity-based feature tracking methods like the method of Lucas and Kanade [8] and Tomasi and Kanade [16] we employ here.

3.2 Focus on Moving Features

The first step towards motion-based feature tracking is to find regions where we expect intensity changes. To do this, we first build the difference image I_{diff} of two temporally adjacent frames $I_{(i)}$ and $I_{(i+1)}$ is build, as can be seen in Fig. 1a) and b):

$$I_{diff} = I_{(i)} - I_{(i+1)}$$

The difference image I_{diff} can be binarized by a static threshold (Fig. 1c) to obtain a first approximation for a mask image I_{mask}, which will define the region-of-interest in which existing features are tracked and new features are sought. Because of small motion variations from one frame to the next one, a fixed threshold of 10% of the mean intensity $\mu_{(I)}$ of the difference image I_{diff} can be applied to the difference image I_{diff} to yield the mask image I_{mask}:

$$\mu_{(I)} = \frac{1}{n}\sum_{x=1}^{n} I(x), \quad n = \text{number of pixels of image } I.$$

$$I_{mask}(x,y) = \begin{cases} 1, \Leftrightarrow I_{diff}(x,y) \geq 0.1 \cdot \mu_{(I_{diff})} \\ 0, \Leftrightarrow I_{diff}(x,y) < 0.1 \cdot \mu_{(I_{diff})} \end{cases}$$

To allow a larger range for the search for features to track, the mask image I_{mask} is treated by the dilation operation δ with a structuring element B (Fig. 1d) as described by Soille [13]:

$$I_{mask} = \delta_B(I_{mask})$$

The final mask image I_{mask} defines the regions where we track features. (Fig. 1e).

3.3 Initialization of New Features

For the definition of features, we used the 'good features to track'-criterion defined by Shi and Tomasi [11] and Bouget [2]. Here, a feature is defined as a window or point, which can be tracked by optimizing some matching criterion with respect to affine transformations. As described in [2], it is important to find an acceptable threshold for the acceptance of new features. If the capture environment has constant light and contrast conditions, it is possible to define just an initial threshold by applying this method to the first image $I_{(0)}$ and use the resulting image $I_{f\,(0)}$ to define an overall threshold of 10% of its intensity:

$$thresh = 0.1 \cdot \mu_{(I_{f\,(0)})}$$

If the environment tends to change, it may be necessary to repeat this procedure to adapt the threshold to actual image conditions. Before the method for the detection of 'easy-to-track'-features can be applied to the image regions defined by the mask image, it is necessary to remove all regions where features already exist. Because already known features are presumably again good candidates and multiple initialization of the same feature has to be avoided, these regions are removed from the mask image. The 'easy-to-track' method can then be applied to the regions defined in the image mask, and new features found in these regions can be added to the existing feature set.

3.4 Rejection of Features

Following the recommendation of [2], a feature is declared 'lost' when it falls outside the image boundaries or a final error function is larger than a predefined threshold; here the latter occurs when the gradient in the feature region tends to zero. Additionally, a feature is declared 'lost' when the recommended new position falls outside the mask boundaries.

4. IMPLEMENTATION

The feature tracking approach presented in this paper is mainly based on the KLT-tracking method described in [8] and [16]. The implementation follows the pyramidal KLT feature tracking implementation by [2].

In a first step, a threshold for the 'easy to track' feature detection is calculated. Then the first difference image is built. The result is binarized by a static threshold and the resulting mask image is dilated with a morphological element in the dimension of the later applied feature. Regions with already know features are removed. In the rest of the image defined by the mask we look for new 'easy-to-track' features and added them to the existing feature set.

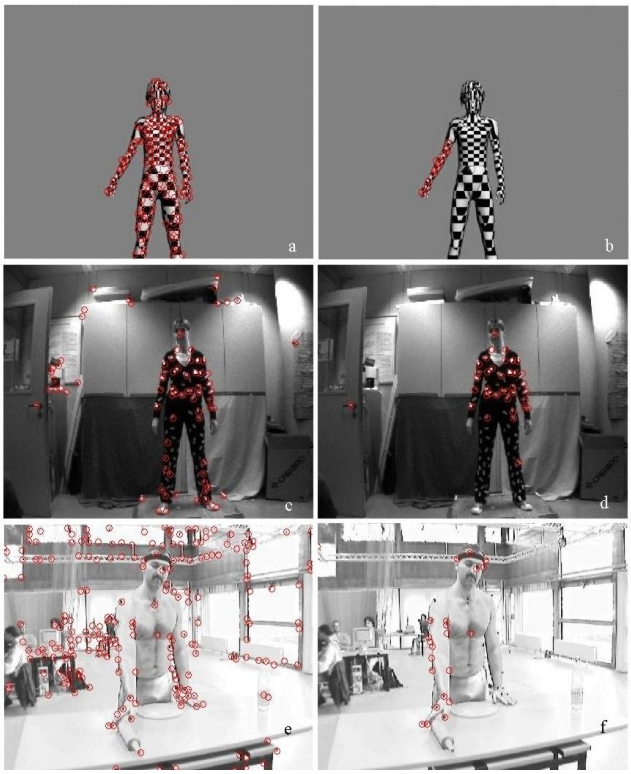

Figure 2: Results of feature reduction with the motion-based feature tracking (right) compared to the general KLT feature tracking (left)

For the tracking a pyramidal approach of the KLT feature tracking method is used with a pyramid level of 2 or 3 and a window size of $2 \cdot 3+1$ pixels. The tracking of a feature in image $I_{(i+1)}$ is based on the detail of its position in image $I_{(i)}$. If a feature violates the predefined acceptance criteria, it is declared lost and removed from the feature set.

5. RESULTS

The algorithm has been evaluated in three video series of which one is an artifical 3D-motion sequence rendered by 3D Studio Max with 16 camears by 30 fps and a resolution of 600x800 px (Fig. 2a,b). The other two series are different video captures of human motion. The first is a motion serie captured by a BumbleBee stereo camera with 20 fps and a resolution of 640x480 px with 12 motion variations with durations in the range of 5 - 20 seconds (Fig 2c,d). The second motion serie has been captured with a 4-camera system with 11 motions and 30 fps, a resolution of 640x480 px and a duration of 15 - 18 seconds per sequence (Fig. 2e,f).

5.1 Feature Correctness

To evaluate the correctness of the tracked features according to a later clustering and definition of body segments, we labeled the markers of over 30 motion videos to define a ground truth for relevant features. Here we only accepted features allocated on moving segments like arms and legs, because they allow a later clustering by segment. The evaluation led to a mean true positive rate of 81.11%, displayed in Fig. 3.

5.2 Feature Stability

To allow clustering based on motion features, it is important to track the features for as long as possible to get comparable motion trajectories. To evaluate this stability, we measured the 'lifetime' of features, namely the number of frames that a feature is tracked until it is rejected. We analyzed the 'lifetime' of features over all three sequences with approximately 50000 features. The mean lifetime over all features was about 100 frames as can be seen in Fig. 4. For the relevant 'lifetime', which must be longer than at least 10 frames we got about 73.77% of all features.

5.3 Runtime Performance

We evaluated the runtime on an Intel Core 2 Duo Processor machine with 2 GB RAM on seven different video sequences with 500 frames per video sequence. The implementation and experiments have been realized with Matlab. As can be seen in Fig. 5 the overall runtime per frame of the presented motion-based feature tracker decreases by 74.54% compared to the usual pyramidal implementation, but depends mainly on the scenario.

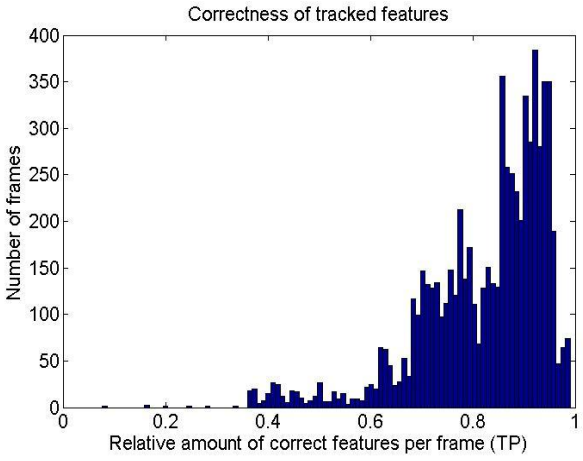

Figure 3: Statistical evaluation of correct detected features per frame with a mean true positive rate of 81.11%.

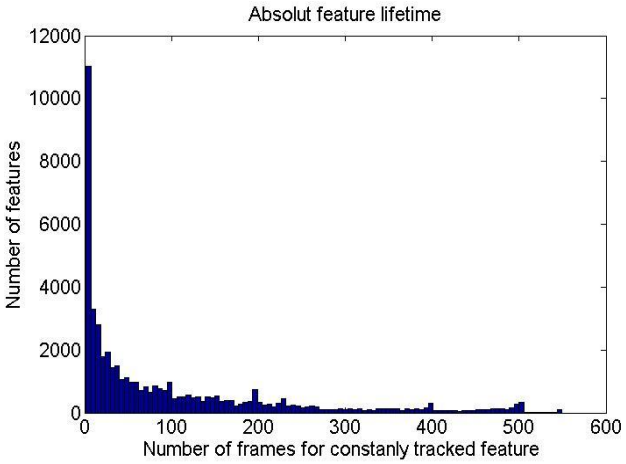

Figure 4: Evaluation of 'lifetime' of features; the mean lifetime over all features is about 100 frames

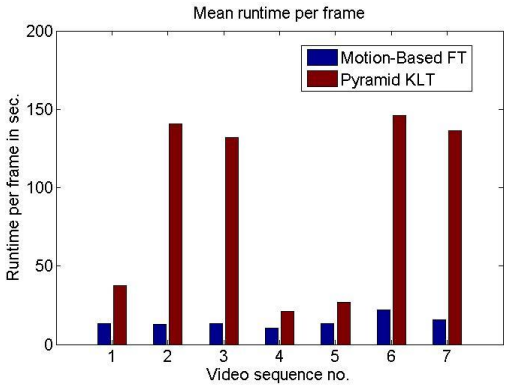

Figure 5: Evaluation of mean runtime per frame of the motion-based feature tracking and the pyramidal KLT

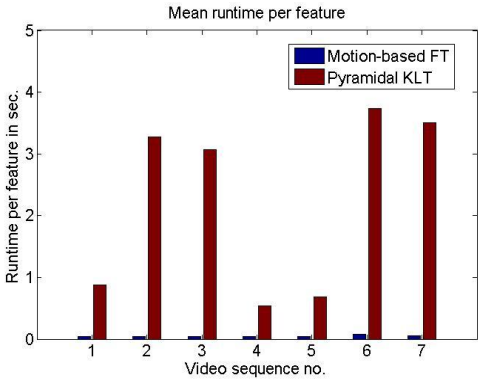

Figure 6: Evaluation of mean runtime per feature of the motion-based feature tracking and the pyramidal KLT

For the second sequence, the runtime decreases by 90.8%, but only about 50.85% for the fourth one.

The runtime performance per feature decreases by 96.50%, shown in Fig. 6. We see that the improved performance depends not only on the decreased number of features to track, but also on the decreased area size, in which new features are searched.

6. CONCLUSION

We have shown an extension of the Kanade-Lucas-Tomasi feature tracking method which allows with a high accuracy the exclusive tracking of moving feature points. This motion-based feature tracking leads to better performance increase, and enjoys a good selection of features relevant for articulated body tracking. The results show potential for clustering feature points according to their 2D motion and the estimation of the underlying body structure.

7. ACKNOWLEDGMENTS

This work was supported by the grant from the Ministry of Science, Research and the Arts of Baden-Württemberg.

8. REFERENCES

[1] Aggarwal, J.K., Cai, Q. 1999. Human Motion Analysis: A Review, Computer Vision and Image Understanding 73, No. 3, March 1999, pp. 428-440.

[2] Bouguet, J.-Y. 2002. Pyramidal implementation of the Lucas Kanade feature tracker, description of the algorithm. Technical report, Intel Corporation.

[3] Cedras, C., Shah, M. 1994. A Survey of Motion Analysis from Moving Light Displays, IEEE Conf. on Computer Vision and Pattern Recognition, 1994, pp. 214-221.

[4] Corazza S., Mündermann L., Andriacchi T., 2007. A framework for the functional identification of joint. centers using markerless motion capture, Validation For The Hip Joint, Journal of Biomechanics, 2007

[5] Giese, M. A., Poggio T. 2003. Neural mechanisms for the recognition of biological movements and action. Nature Reviews Neuroscience 4, 2003, pp. 179-192.

[6] Johansson, G., 1973. Visual perception of biological motion and a model for its analysis, Perception & Psychophysics, Vol. 14, No. 2, pp. 201 - 211.

[7] Li, B.H., Meng, Q.G., Holstein, H., 2008. Articulated motion reconstruction from feature points, Pattern Recognition, Vol. 41, No. 1, January 2008, pp. 418 - 431

[8] Lucas, B.D., and Kanade, T., 1981. An Iterative Image Registration Technique with an Application to Stereo Vision, Proceedings of the 7th Int. Joint Conf. on Artificial Intelligence 1981, pp. 121 – 130.

[9] Moeslund, T.B., Granum, E., 2001. A survey of computer vision-based human motion capture, Computer Vision and Image Understanding, Vol. 81 , No. 3, March 2001, pp. 231 – 268.

[10] Moeslund, T.B., Hilton. A., Krüger, V., 2006. A survey of advances in vision-based human motion capture and analysis, Computer Vision and Image Understanding, Vol. 104 , No. 2, November 2006, pp. 90 – 126.

[11] Shi, J., Tomasi, C. 1994. Good Features to Track, 1994 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'94), 1994, pp. 593 - 600.

[12] Silaghi, M.-C., Plänkers, R., Boulic, R., Fua, P., Thalmann, D., 1998, Local and Global Skeleton Fitting Techniques for Optical Motion Capture, LNCS, Vol. 1537, Proc. of the International Workshop on Modelling and Motion Capture Techniques for Virtual Environments, pp. 26 – 40.

[13] Soille, P. 2003 Morphological Image Analysis: Principles and Applications. 2. Springer-Verlag New York, Inc.

[14] Song, Y., Goncalves, L., Di Bernardo, E., Perona,P., 1999. Monocular Perception of Biological Motion - Detection and Labeling, Proc. of the Int. Conf. on Computer Vision, Vol. 2 , pp. 805 – 812.

[15] Song, Y., Goncalves, L., Perona, P., 2003. Unsupervised Learning of Human Motion, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25 , No. 7, July 2003, pp. 814 – 827

[16] Tomasi, C. and Kanade, T. 1991. Detection and tracking of point features. Technical Report CMU-CS-91-132, School of Computer Science, Carnegie Mellon University, April 1991

[17] Zhang, X., Lee, S.-W., Braido, P., 2004. Towards an integrated high-fidelity linkage representation of the human skeletal system based on surface measurement, International Journal of Industrial Ergonomics, Vol. 33, No. 3, March 2004, pp. 215-227.