# Unified Target Detection and Tracking Using Motion Coherence

Markus Enzweiler* [1,2] Richard P. Wildes [1] Rainer Herpers [1,2]

[1] Dept. of Computer Science, York University, Toronto, ON, Canada, M3J 1P3
[2] Dept. of Comp. Sci., Bonn-Rhein-Sieg Univ. of Applied Sciences, D-53757 St. Augustin, Germany
{menz, wildes, herpers}@cs.yorku.ca

## Abstract

*This paper presents a unified approach to adaptive target detection and tracking. The unifying concept is "coherent motion energy", a measure of the extent to which a single motion dominates local spatiotemporal structure. There are three major components to the approach. First, a multiresolution analysis of coherent motion energy is used to detect salient dynamic targets. Second, a robust affine transformation estimator is used to recover frame-to-frame target motion across regions of interest defined by coherent motion. Third, a method of template adaptation based on coherent motion weighted goodness of match is used to drive automatic template update. Empirical evaluation of the approach shows the contribution of the various components and documents strong performance of the integrated whole.*

## 1. Introduction

The perception and interpretation of motion provides a basic mechanism for guiding action. While humans are good at detecting and tracking targets from moving backgrounds, machine vision approaches usually provide satisfactory results only under well-defined assumptions. The detection and tracking of biological targets is particularly challenging, as natural shapes change non-rigidly over time.

The problem of recovering the motion of potential targets from image sequences has been widely addressed [4, 22]. General approaches to estimating target trajectories have been both token (e.g., [16, 24, 25]) and area (e.g., [2, 9, 18]) based. Geometric deformation of tracked regions has been addressed to allow for a degree of 2D non-rigidity (e.g., [10, 19, 26]). Other work has computed motion parameters for connected components of points, supported by geometric and kinematic filtering [21]. Recently, several additional lines of research have been concerned with non-rigid 3D shape recovery and tracking [7, 27, 30]. Still other work has concentrated on adaptive techniques to accommodate template variation across time [12, 17, 20].

Most closely related to the current contribution is previous work that has employed spatiotemporal, directionally selective filters to detect salient targets of interest [1, 28]. Especially in terms of disregarding a scintillating or oscillating background, such techniques have proven to be more efficient than change detection based on temporal differences [28]. Such filters also have been applied to make qualitative distinctions between different patterns of motion in terms of oriented energy signatures [29]. Of particular interest in the current work is the energy signature of coherent motion, where a single trajectory dominates a local region in space-time, as a cue to detect potential targets.

In the light of previous research, the main contributions of the current approach are as follows. A novel adaptive algorithm to detect and track multiple, non-rigidly moving objects over time is presented. There are three components to the algorithm. (i) A measure of coherent motion is used to detect dynamic targets. (ii) A robust affine transformation estimator is used to recover frame-to-frame target motion across coherent motion defined regions of interest. (iii) An adaptive method, based on coherent motion weighted goodness of match, is used to drive template update. A systematic empirical evaluation quantitatively documents the contribution of each of the algorithm's components. While components of the described approach have been considered in previous research, their unified synthesis via a measure of coherent motion is novel as is the empirical delineation of how each component contributes to the whole. Further, in contrast to most extant approaches, the proposed approach supports both automatic initialization (detection) and generation of correspondences (tracking) within a single framework.

## 2. Technical approach

### 2.1. Coherent motion energy

In general, the motion of non-rigid objects in image sequences results from the projection of the three-dimensional

*Currently at the Faculty of Comp. Sci., University of Ulm, D-89069 Ulm, Germany, markus.enzweiler@informatik.uni-ulm.de.

Proceedings of the IEEE Workshop on Motion and Video Computing (WACV/MOTION'05)
0-7695-2271-8/05 $ 20.00 IEEE

object motion onto the two-dimensional image plane. Under the assumption that natural targets exhibit a certain pattern of texture, coherent motion (i.e., translation) generates locally linear structure in the spatiotemporal domain. To detect such spatiotemporal gradients, an oriented energy representation of the image sequence is used, as follows [1, 29].

Measures of oriented energy, $E_R$, $E_L$, $E_U$, $E_D$, corresponding to rightward, leftward, upward and downward motion are extracted by pointwise rectification and summation of the responses of a quadrature pair of orientation selective bandpass filters at four orientations along spatiotemporal diagonals indicative of rightward, leftward, upward and downward motion. Here, a filter pair consisting of broadly tuned separable and steerable filters based on the 3D second derivative of a Gaussian, $G_2$, and their corresponding Hilbert transforms, $H_2$, are employed [11]. Given that the filters combine selection for scale and orientation, the filtering operation is extended to $n$ scales to efficiently cover different frequency bands with respect to orientation via an oriented Gaussian pyramid decomposition [11].

Following [29], a spatiotemporal region corresponding to coherent motion is characterized by the ratio of the difference between two opponent energy measures and their sum. Thus, measures of coherent motion energy can be captured both horizontally ($E_R$, $E_L$) and vertically ($E_U$, $E_D$) as a function of time, within the frequency band extracted by the quadrature filter pair:

$$E_{hor} = \left|\frac{E_R - E_L}{E_R + E_L + \epsilon}\right|, \qquad E_{ver} = \left|\frac{E_U - E_D}{E_U + E_D + \epsilon}\right| \quad (1)$$

A small bias $\epsilon$ (approx. 1 % of the maximum energy) is added for stability in case of overall low energy. A pixelwise maximum operation is used to combine $E_{hor}$ and $E_{ver}$ into a single measure of coherent motion energy, $E_{mot}$, where regions with high values are indicative of potential coherently moving targets.

## 2.2. Multiscale target detection

Candidates for coherently moving targets, i.e., regions with high coherent motion energy, are not necessarily caused by actual moving objects. A noisy, scintillating or oscillating background might also effect a peak in the coherent motion energy signal, especially as slow noisy background motion might be coherent over a small temporal interval. Under the assumption that most targets of interest are composed of a wide range of spatiotemporal frequencies (i.e., textured objects in motion), a candidate region corresponding to a coherently moving object exhibits a significant amount of energy across several scales. Since many common noise sources concentrate their energy in relatively high frequencies, aberrations in the coherent motion energy signal due to noise or noisy backgrounds generally do not persist in the coarser scales of a multiscale bandpass representation.

A coarse-to-fine segmentation strategy [2, 3], based on radial scanline clustering [23], is employed across $n$ levels of the multiscale representation to extract regions with high coherent motion energy. On a finer scale, a region exhibiting coherent motion energy is discarded, if it does not have a corresponding parent on a coarser scale with respect to both size and spatial location. The multiscale extension of the radial scanline clustering provides additional robustness against high-frequency noise and large interframe displacements. An accepted region of high motion energy, referred to as $E_i$ in the remainder, is thereby represented by the bounding box and the center position $(\widetilde{l_x^i}, \widetilde{l_y^i})$, which is used as an initial estimate for the location of a salient target.

## 2.3. Target representation

Since the current work is mainly concerned with the recovery of the motion of multiple non-rigid targets, the tracking algorithm relies on the photometric structure of the multiple tracked objects. In doing so, targets exhibiting a similar geometric structure, e.g. two walking persons, can still be distinguished based on their spatial appearance.

Given the a priori information about the location of coherently moving targets, $(\widetilde{l_x^i}, \widetilde{l_y^i})$, defined as the center of salient regions, $E_i$, a dynamic internal template representation is automatically initialized from the first frame of an image sequence for each target [15]. An internal template, $T_i$, initially is defined by the image region specified by the bounding box of the extracted salient target region, $E_i$, and is dynamically matched to each frame of the sequence according to a match criterion based on the brightness constancy assumption: The brightness (image intensity) values of each target remain (approximately) constant between subsequent frames in a temporal sequence of images [13].

## 2.4. Robust motion estimation

To match the template $T_i$ to the data, the motion of the target represented by $T_i$ is estimated and compensated for, as defined by a recovered flow estimate. Under the assumption that the variation in distance within a target is small compared to the target-to-camera distance, a parametric affine motion model is used to constrain flow recovery (as used to good advantage in previous target tracking research, e.g., [19, 21, 26]) and embedded in an hierarchical and robust estimation framework [4, 5, 6]. In a parametric model, the optical flow constraint equation is given by

$$\nabla^\top I \vec{u}(\vec{a}) + I_t = 0, \quad (2)$$

where $\nabla^\top I = (I_x, I_y)$ and $I_t$ denote the first-order partial derivatives of the local brightness structure, $\vec{u} = (u, v)^T$ represents the flow vector and $\vec{a}$ is a set of model parameters, specifying the motion of the local region. In case of an

affine motion model, $u$ and $v$ are defined as:

$$u(x,y) = a_0 + a_1 x + a_2 y \tag{3}$$

$$v(x,y) = a_3 + a_4 x + a_5 y \tag{4}$$

The recovery of the affine parameters $\vec{a} = (a_0, a_1, \ldots, a_5)^\top$ is formulated as the minimization of an energy-weighted error measure, $\xi(\vec{a})$, over coherent energy defined salient regions $E_i$. With respect to the possible deviation of the target motion from the affine model, i.e. under non-rigid and articulated motion, and the influence of non-target-pixels due to the broad response characteristics of the oriented energy operator [11], a standard least-squares solution might be error-prone. Thus, the minimization problem is reformulated within a robust estimation framework, employing an error norm $\rho(\eta, \sigma)$, to diminish the influence of outliers [6], with weights $0 \leq w = w(x,y) = E_{mot}(x,y) \leq 1$, used to make a point's contribution proportional to its motion coherence:

$$\min_{\vec{a}} \xi(\vec{a}) = \min_{\vec{a}} \sum_{(x,y)\in E_i} w\rho(\nabla^\top I \vec{u}(\vec{a}) + I_t, \sigma) \tag{5}$$

To facilitate a smooth transition between inliers and outliers, the Geman-McClure error norm [14] is employed in the current implementation, as suggested in [6].

To recover motion with larger displacement, the minimization algorithm is based on a coarse-to-fine gradient descent technique. At each level of a Gaussian pyramid, the image at the corresponding level is warped according to the previous estimate of the affine parameters. The warped image is then used to compute a residual change in the parameters according to the gradient descent scheme. The updated affine parameters are used as initial estimates at the next pyramid level, until the finest level is reached. Implementation details of the estimator are given in [5, 6].

As formulated, the robust estimator is still prone to instabilities resulting from the general aperture problem [4, 6]: If the local neighborhood is too small or devoid of discernable structure, the motion is not adequately constrained; on the other hand, overly large regions of spatial support can be inconsistent with the affine motion model. To ameliorate such difficulties, a maximum deviation of the recovered affine parameters from the identity transform is enforced. This constraint is realized as a threshold on the Frobenius matrix norm applied to the difference of the first-order affine parameter matrix and the identity matrix. If the threshold is exceeded, then the transformation is restricted to that captured by the translational components only; otherwise, the full affine transformation is used to characterize the motion. In this regard, the algorithm has an automatic procedure for selecting the order of the motion model that is applicable to the data at hand.

The recovered affine parameters that align a target template, $T_i$, with a region, $R_i$, in an image frame, $I$, refine the coherent energy positional estimate associated with $I$, $(\widetilde{l}_x^i, \widetilde{l}_y^i)$, to a location with subpixel precision, $(l_x^i, l_y^i)$, to be associated with the target in the given frame. The temporally sorted set of locations across the sequence comprise the trajectory of the target. Additionally, the estimated affine parameters associated with each frame-to-frame transition are available to augment the state vectors associated with the target.

### 2.5. Confidence measure

To evaluate the goodness of each target location, $(l_x^i, l_y^i)$, recovered by the affine motion estimator (Sec. 2.4), a normalized confidence measure $0 \leq \pi(\vec{a}) \leq 1$ is computed along with each template alignment. For consistency with the approach to motion estimation, confidence is based on the residual error, $\xi_{res}(\vec{a})$, associated with the recovered affine parameter vector, $\vec{a}$,

$$\xi_{res}(\vec{a}) = \sum_{(x,y)\in E_i} w\rho(\nabla^\top I \vec{u}(\vec{a}) + I_t, \sigma), \tag{6}$$

with $\rho(\eta, \sigma)$ the Geman-McClure error norm, the area of summation for the region of interest, $E_i$, defined by coherent energy and $0 \leq w \leq 1$ the associated weighting coefficients, exactly as for motion estimation purposes, c.f., Eq. (5).

Toward the definition of a normalized confidence measure, the residual error, $\xi_{res}(\vec{a})$, is itself normalized by recalling that the Geman-McClure error norm, $\rho(\eta, \sigma) = \eta^2/(\eta^2 + \sigma^2)$, is by definition bounded from above by 1 (and from below by 0). Correspondingly, the residual error, $\xi_{res}(\vec{a})$, is bounded from above by $\sum_{(x,y)\in E_i} w(x,y)$, the sum of the weighting coefficients over the region of interest $E_i$ (and bounded from below by 0). Pulling these observations together, the residual error, $\xi_{res}(\vec{a})$, can be normalized through division by $\sum_{(x,y)\in E_i} w(x,y)$. Finally, a normalized confidence measure, $0 \leq \pi(\vec{a}) \leq 1$, is given by

$$\pi(\vec{a}) = 1 - \frac{\xi_{res}(\vec{a})}{\sum_{(x,y)\in E_i} w(x,y)} \tag{7}$$

with confidence increasing as the value approaches 1.

### 2.6. Template adaptation

During tracking, a low match confidence, $\pi(\vec{a})$, is taken to indicate a change in the tracked target. The current state of the internal template, $T_i$, corresponding to the target does not correctly reflect the actual data. Possible causes for change in the tracked object, especially regarding non-rigid targets, are non-affine shape deformations, perspective distortions and change in the photometry of the object, i.e. varying illumination. To compensate for deviation of the

template, $T_i$, from the actual data, a temporal integration technique is employed to dynamically update $T_i$ [15].

The confidence measure $\pi(\vec{a})$ is used as a weighting coefficient, directly controlling the amount of temporal integration applied to the template $T_i$. $\pi(\vec{a})$ is inversely proportional to the degree of template adaptation, i.e., the lower the match confidence, the more update is necessary to prevent the algorithm from losing track of the target, resulting in:

$$\hat{T}_i = \pi(\vec{a})T_i + (1 - \pi(\vec{a}))R_i \tag{8}$$

The updated internal template $\hat{T}_i$ is calculated as the confidence-weighted sum of the template $T_i$ and the corresponding image region $R_i$, which was aligned with the template by the affine estimator. As a result, the adapted internal template $T_i$ used to track the corresponding target in frame $I(x, y, t)$ captures both the geometric change of the target in frame $I(x, y, t)$, by means of the image registration operation (see Sec. 2.4), and the change in the photometric structure of the target in the previous frame $I(x, y, t - 1)$, estimated by the confidence value $\pi(\vec{a})$ of the alignment of the template $T_i$ to the data.

## 3. Empirical evaluation

The described approach to target detection and tracking has been implemented in software and tested on an illustrative set of synthetic and natural image sequences, by means of computing an RMS error measure $\xi_{RMS}$ between the recovered target locations $(l_x^i, l_y^i)$ and ground truth for each frame.

To evaluate performance against exact ground truth and systematically manipulated signal-to-noise ratio (SNR), an image sequence of a fish translating through an aquarium has been synthetically generated and corrupted with additive, zero mean Gaussian noise of varying standard deviation. Fig. 1 (top left) shows four frames of this sequence at signal-to-noise ratios of $\varrho = \infty$, $\varrho = 100$, $\varrho = 50$ and $\varrho = 20$ (left to right, top to bottom) as well as recovered vs. ground truth trajectories.

The addition of Gaussian noise represents a violation of a basic premise of the proposed approach, the brightness constancy assumption, manifesting in an increase of the residual error, $\xi_{res}(\vec{a})$, of the recovered affine transform, as the standard deviation of the noise increases. Consequently, the RMS error $\xi_{RMS}$, as shown in Fig. 2 (first group), increases as function of additional Gaussian noise, since the recovered affine parameters $\vec{a}$ become less ideal. Given that the confidence $\pi(\vec{a})$ (see Sec. 2.5) directly depends on the residual error $\xi_{res}(\vec{a})$, the confidence of each recovered target location $(l_x^i, l_y^i)$, that is the degree to which this position can be trusted, drops, as the noise increases. The characteristic of the confidence measure as a function of frame number is shown in Fig. 3.

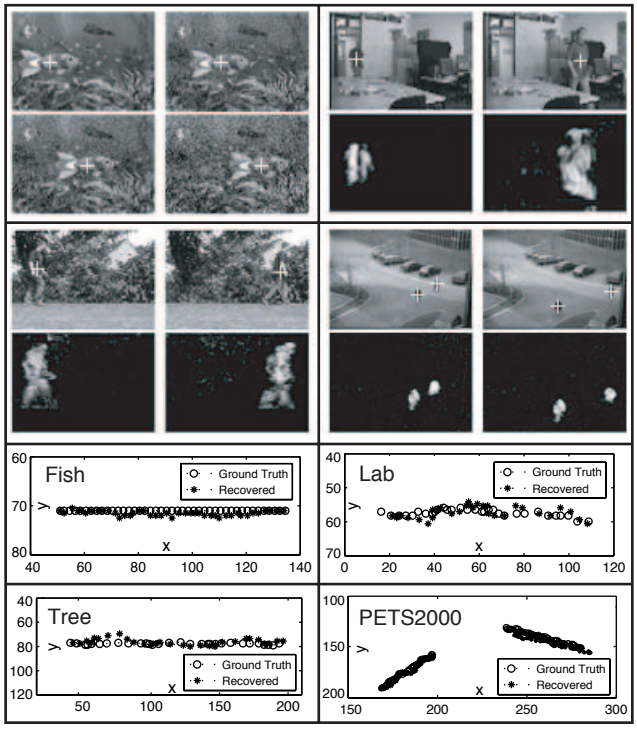


**Figure 1. Left to right, top to bottom: Four frames of the *Fish* sequence $(x \times y \times t{=}240 \times 180 \times 80)$ at different SNR $\varrho$ (see Sec. 3), two frames each of the *Lab* $(x \times y \times t{=}240 \times 180 \times 65)$, *Tree* $(x \times y \times t{=}240 \times 180 \times 60)$ and *PETS2000* $(x \times y \times t{=}320 \times 240 \times 70)$ sequences, along with corresponding coherent energy maps and marked recovered target locations, (+). The lower plots show recovered vs. ground truth trajectories.**

It is important to note, that even at a low signal-to-noise ratio of $\varrho = 20$, the tracking algorithm performs reasonably well, with an RMS error of $\xi_{RMS} = 6.15$ and an average confidence value across the sequence of $\pi(\vec{a}) = 0.8527$. The reasons for the observed performance are twofold: First, the target of interest does not deform non-rigidly during the sequence, which restricts the affine estimator to translational components only. Second, the rich texture on the target contributes to a strong coherent motion energy signal (see Sec. 2.1), which still allows for a successful extraction of the coherently moving object in combination with the multiscale target detection algorithm, even at low signal-to-noise ratios.

To document the contribution of various components of the approach, results are shown with respect to three natural image sequences and hand-picked ground truth as major algorithmic components are incorporated systematically. First, results are shown for a standard normalized correlation tracker, restricted to an $n \times n$ window ($n = 15$) about the predicted target location, based on linear extrapolation of target speed from previous estimates (algorithm

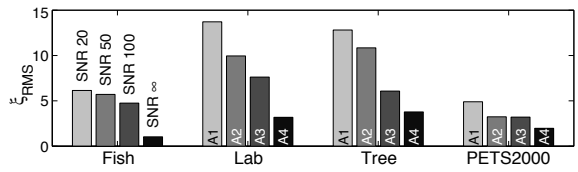


**Figure 2. RMS error for *Fish* ($\varrho \in [20, 50, 100, \infty]$), *Lab*, *Tree* and *PETS2000* ($A_i, i \in [1, 2, 3, 4]$).**

$A_1$). Second, results are shown for the same correlation tracker, but now with search about a location indicated by coherent motion energy, as described in Sec. 2.1 (algorithm $A_2$). Third, results are shown for correlation tracking, constrained by coherent motion energy, but further augmented to include template adaptation, as shown in Sec. 2.6 (algorithm $A_3$). Fourth, results for the complete algorithm are shown, as the affine estimator (as described in Sec. 2.4) replaces the correlation estimator, while the motion energy and adaptive template components remain in place (algorithm $A_4$). For all versions, the multiscale motion energy detection algorithm is used for automatic initialization (see Sec. 2.2). In all correlation-based algorithms the proposed confidence measure $\pi(\vec{a})$ dependent on the residual error of the affine transform estimation is replaced by a similarity measure independent from photometric effects based on the normalized correlation coefficient (e.g., [8]). Results are shown in Figs. 1, 2 and 3.

The *Lab* sequence shows a person walking through an indoor environment. There are significant amounts of depth motion, as well as non-rigid deformations. A sequence from the *PETS2000* dataset[1] is used to evaluate the proposed approach in terms of multiple small targets and large target-to-camera distance. The influence of noisy scintillating background motion, i.e., fluttering leaves, as an example of structured noise, in combination with a non-rigidly moving target, is evaluated by means of the *Tree* sequence. Note that in this sequence attempts to constrain tracking based on temporal differencing would fail, as there is temporal change over large portions of the background [28].

The comparison of the RMS error measure (Fig. 2) for all trackers under consideration shows that each proposed component to the algorithm contributes to the decrease of the tracking error. The delineation of targets of interest based on motion coherence ($A_2$) is clearly superior to using estimated target speed ($A_1$) as a correlation search constraint, since it provides a strong means of parsing the data into relevant and irrelevant portions. The addition of temporal integration ($A_3$) as a way to adapt the template to changes in the data (see Sec. 2.6) proves useful especially in sequences, where the target's geometry is constantly changing, i.e., in the *Lab* and *Tree* sequences. Template adaptation is only a minor improvement, if the target's geometry and photometry remains approximately constant over the sequence, i.e., as seen in the *PETS2000* sequence.

[1] `http://peipa.essex.ac.uk/ipa/pix/pets/`

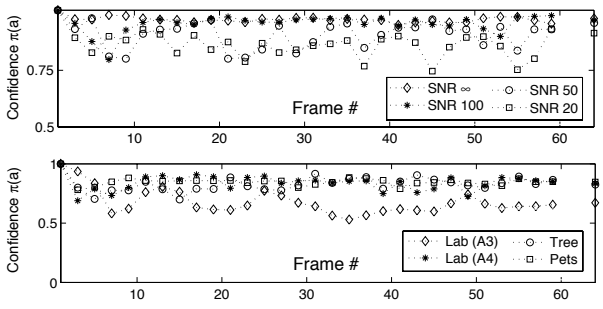


**Figure 3. Confidence $\pi(\vec{a})$ as a function of frame number for *Fish* ($A_4$) (top) and *Lab* ($A_3$ and $A_4$), *Tree* ($A_4$) and *PETS2000* ($A_4$) (bottom). The mean confidence over all frames is shown at the right of the plot.**

Temporal integration without adequate image registration leads to the problem that a significant number of frames might need to be considered before the template has fully adapted to a change in the target, since both variations in the geometry and photometry must be compensated. This behavior is evident in the evolution of the confidence measure $\pi(\vec{a})$ of the *Lab* sequence, which exhibits an oscillating form, as seen in Fig. 3 (algorithm $A_3$). Replacing the correlation estimator with the affine motion estimator ($A_4$) solves this problem, since the geometric change in the target is accounted for by the recovered affine transform; whereas, temporal integration is employed only to handle the photometric and non-affine variations in the target.

It also is of note that non-affine shape deformations, i.e., due to significantly non-rigidly moving objects (e.g., a person walking parallel to the image plane, as in the *Tree* sequence) can cause jitter in the confidence values and lead to a higher RMS error, since the target motion deviates from the affine model. In this regard, the importance of the unifying concept of motion coherence becomes apparent. Since the presented approach continuously uses coherent motion energy to define what part of the data should be incorporated in the calculation, it inherently focuses computation on linearly moving regions of the target, which exhibit a stronger coherent energy signal, i.e., the energy maps of the *Tree* sequence highlight the upper part of body while attenuating the influence of the incoherent motion of the legs, as shown in Fig. 1. By adhering to this framework, the algorithm manages to perform reliably in many situations, without losing track of the targets of interest, even if there is a significant deviation from the adopted affine motion parametrization.

On the whole, the proposed approach represents a marked improvement over correlation-based trackers with respect to lower RMS error and higher confidence values

(Figs. 2 and 3). Improvement is observed for a variety of scenarios, including small targets (*PETS2000*), noisy backgrounds (*Tree*) and non-rigid targets (*Lab* and *Tree*). Further, while many extant approaches rely on manual initialization, all results presented here are auto-initialized through the integrated target detection procedure (Sec. 2.2).

## 4. Summary

The decomposition of an image sequence into the trajectories of salient targets supports subsequent processing in terms of interpretation, identification and classification. Toward such ends, this paper has presented an approach to detecting and tracking non-rigid objects in image sequences. The approach employs an algorithm consisting of a multiscale target detector and a constrained, adaptive tracker with subpixel precision. Coherent motion energy plays a prominent role in the definition and unification of the major components of the approach. Empirical evaluation of a software realization of the approach shows that it is able to recover the projected two-dimensional motion of non-rigid targets at various spatial scales with an interesting level of robustness to noise.

## Acknowledgements

*M. Enzweiler acknowledges the support of the German National Merit Foundation and the German Academic Exchange Service (DAAD). Portions of this research were supported by an NSERC grant to R. P. Wildes. R. Herpers acknowledges the support of the Bundesministerium für Bildung und Forschung (BMBF-IB), grant WTZ CAN 02/011.*

## References

[1] E. H. Adelson and J. R. Bergen. Spatiotemporal energy models for the perception of motion. *JOSA A*, 2(2):284–299, 1985.

[2] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *IJCV*, 2(3):283–310, 1989.

[3] C. H. Anderson et al. Change detection and tracking using pyramid transform techniques. In *Proc. SPIE IRCV*, pages 72–78, 1985.

[4] S. S. Beauchemin and J. L. Barron. The computation of optical flow. *ACM CS*, 27(3):433–467, 1995.

[5] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *Proc. ECCV*, pages 237–252, 1992.

[6] M. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *CVIU*, 63:75–104, 1996.

[7] M. Brand and R. Bhotika. Flexible flow for 3D non-rigid tracking and shape recovery. In *Proc. CVPR*, pages 315–322, 2001.

[8] M. Z. Brown, D. Burschka, and G. D. Hager. Advances in computational stereo. *IEEE PAMI*, 25(8):993–1008, 2003.

[9] P. Burt, C. Yen, and X. Xu. Local correlation measures for motion analysis: A comparative study. In *Proc. PRIP*, pages 269–274, 1982.

[10] W. Förstner. Reliability analysis of parameter estimation in linear models with application to mensuration problems in computer vision. *CVGIP*, 40:273–310, 1987.

[11] W. T. Freeman and E. H. Adelson. The design and use of steerable filters. *IEEE PAMI*, 13(9):891–906, 1991.

[12] G. D. Hager and P. N. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *IEEE PAMI*, 20(10):1025–1039, 1998.

[13] B. K. P. Horn. *Robot Vision*. MIT press, 1986.

[14] P. J. Huber. *Robust Statistical Procedures*. SIAM Press, 1977.

[15] M. Irani, B. Rousso, and S. Peleg. Computing occluding and transparent motions. *IJCV*, 12(1):5–16, 1994.

[16] O. Javed, Z. Rasheed, K. Shafique, and M. Shah. Tracking in multiple cameras with disjoint views. In *Proc. ICCV*, pages 952–957, 2003.

[17] A. Jepson, D. Fleet, and T. El-Maraghi. Robust online appearance models for visual tracking. *IEEE PAMI*, 25(10):1296–1311, 2003.

[18] B. D. Lucas and T. Kanada. An iterative image registration technique with an application to stereo vision. In *Proc. DARPA IUW*, pages 121–130, 1981.

[19] R. Manmatha and J. Oliensis. Extracting affine deformations from image patches. In *Proc. ICPR*, pages 754–755, 1993.

[20] I. Matthews, T. Ishikawa, and S. Baker. The template update problem. *IEEE PAMI*, 26(6):810–815, 2004.

[21] F. Meyer and P. Bouthemy. Region-based tracking in an image sequence. In *Proc. ECCV*, pages 476–484, 1992.

[22] H. H. Nagel. Image sequence evaluation: 30 years and still going strong. In *Proc. ICPR*, pages 149–158, 2000.

[23] R. Herpers et al. SAVI: An actively controlled teleconferencing system. *IVC*, 19(11):793–804, 2001.

[24] I. Sethi and R. Jain. Finding trajectories of feature points in a monocular image sequence. *IEEE PAMI*, 9(1):56–73, 1987.

[25] M. Shah, K. Rangarajan, and P. S. Tsai. Motion trajectories. *IEEE SMC*, 23(4):1138–1150, 1993.

[26] J. Shi and C. Tomasi. Good features to track. In *Proc. ICPR*, pages 593–600, 1994.

[27] L. Torresani et al. Tracking and modeling non-rigid objects with rank constraints. In *Proc. CVPR*, pages 493–500, 2001.

[28] R. P. Wildes. A measure of motion salience for surveillance applications. In *Proc. ICIP*, pages 183–187, 1998.

[29] R. P. Wildes and J. R. Bergen. Qualitative spatiotemporal analysis using an oriented energy representation. In *Proc. ECCV*, pages 768–784, 2000.

[30] C. R. Wren et al. Real-time tracking of the human body. *IEEE PAMI*, 19(7):780–785, 1997.