Robust visual tracking by coupling 2D motion and 3D pose estimation

Éric Marchand\textsuperscript{1}, Patrick Bouthemy\textsuperscript{1}, François Chaumette\textsuperscript{1}, Valérie Moreau\textsuperscript{2}

\textsuperscript{1}IRISA / INRIA Rennes
Campus de Beaulieu
35042 Rennes Cedex, France

\textsuperscript{2}EDF - Pôle Industrie -
Division Recherche et Développement
6, Quai Watier
78401 Chatou Cedex, France

Abstract

We present an original method for tracking in an image sequence complex objects which can be modeled approximately by a polyhedral shape. The approach relies on the estimation of the object image motion as well as the computation of the object pose. The proposed method fulfills real-time constraints along with reliability and robustness requirements.

1 Introduction

Introduction of the method, its technical details and experiments are presented in this section.

2 Motion-based tracking

The motion of the object is estimated using a suitably chosen motion model. The model is then used to predict the object's future position and orientation. The predicted position is then compared with the actual position of the object in the next frame of the sequence. The difference between the predicted and actual positions is used to update the parameters of the motion model, which is then used to predict the object's position and orientation in the next frame. This process is repeated for each frame of the sequence.

Affine transformation model. Let $X' = [X, X']$, $c$, $v$, $f$ be the parameters of the motion model. Then, the motion of the object is given by $X' = f(t, X, c, v, f)$. The parameters are estimated using a least squares method, which minimizes the difference between the predicted and actual positions of the object.

The method is tested on a variety of real-world sequences, and the results are compared with those of existing methods. The results show that the proposed method is more accurate and robust than the existing methods.
$f^t \in \mathcal{F}$, $b \in \mathcal{B}$, $c \in \mathcal{C}$, $u \in \mathcal{U}$, $ct \in \mathcal{C}t$. $\text{Inf}^t \in \text{tuc}$

$X^{t+1} = \Psi_0(\lambda^{t,N})$ (N)

$2 \text{nic} \Psi_0 \in \text{tr} \in \mathcal{V} \in \mathcal{W}$

$X^{t+1} = \Psi_0(\lambda^{t,N})$ (N)

$2 \text{tn} \Theta = (a_1, a_2, a_3, T_x, T_y, N_x, X^{t+1} = (x, y, z, u, v)(N)$

$W(X^N) = \begin{bmatrix} x & y & z & u & v \end{bmatrix}$

$\text{Inf}^t \text{tr} \text{tuc} \text{ct} \text{dt} \text{ct} \text{ct} \text{ct} \text{ct} \text{ct}$

$\text{Affine transformation estimation}$

$\text{Model-based tracking}$

$X^{t+1} = \Psi_0(\lambda^{t,N})$ (N)

$2 \text{nic} \Theta' = \Theta - \mathcal{V} \Phi_0 \text{tr} \text{tuc}$

Computing normal displacements. $\mathcal{V} \Phi_0 \text{tr} \text{tuc}$

$\text{Inf}^t \text{tr} \text{tuc} \text{ct} \text{ct} \text{ct} \text{ct} \text{ct} \text{ct}$

$\text{Inf}^t \text{tr} \text{tuc} \text{ct} \text{ct} \text{ct} \text{ct} \text{ct} \text{ct}$

$X^{t+1} = \Psi_0(\lambda^{t,N})$ (N)

$2 \text{nic} \Theta' = \Theta - \mathcal{V} \Phi_0$
Experimental results

Conclusion

5. Conclusion

The experiments conducted in this study have shed light on the perfor-
mance of the proposed algorithm in various practical scenarios. The
demonstrated accuracy and efficiency make it a promising approach
for real-world applications, especially in the domain of medical
imaging and computer vision.

Evaluating the performance of the proposed method, we find that
it excels in terms of speed and accuracy, outperforming existing
methods in several test cases. The robustness of the algorithm
against noise and variations in input data is a significant advan-
tage, making it suitable for a wide range of applications.

In summary, the proposed method offers a significant leap in
technological advancement, paving the way for further research
and development in the field of image processing.

References

[1] X. Zhang, Y. Wang, and Z. Zhang, “Deep Learning for Medical Imag-
1–12, 2016.

Y. W. Sun, T. Dollar, H. Adam, and R. S.dorf, “Mobilenets: In-
novative, Efficient Convolutional Networks for Mobile Vision


Image Recognition,” in Proceedings of the IEEE Conference on

Tell: A Neural Image Caption Generator,” in Proceedings of the
**Footic** : Nut tracking: (a) tracking with only 2D motion estimation (b) tracking with both 2D motion estimation and 3D pose computation

\[
\mathbf{u}^b = \mathbf{u}^r_t - \mathbf{t} \mathbf{r}^t_w = \mathbf{r}
\]

\[E^2 = \mathbf{F} \cdot (\mathbf{P}^k - \mathbf{b}) \cdot \mathbf{R} \cdot \mathbf{c} = \mathbf{d}_c + \mathbf{b}_t \cdot \mathbf{t}^N \cdot \mathbf{b}_f \cdot \mathbf{i}^t.
\]

References


**Frui**c2: Box tracking: Distortion is very important due to a 3.5mm lens

**Frui**c 5: For different experiments (a, b, c, d) nut tracking featuring various difficulties (see text for details), quite satisfactory results are obtained (three distinct instants of the sequence are only displayed for each example).