J. Vis. Commun. Image R. 22 (2011) 465–478

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Heuristic algorithm for visual tracking of deformable objects ☆

Elena Sánchez-Nielsen [a,*], Mario Hernández-Tejera [b]

[a] Department of Statistics, Operations Research and Computer Science, University of La Laguna, 38271 La Laguna, Spain
[b] Institute of Intelligent Systems and Numerical Applications in Engineering, University Campus of Tafira, 35017 Gran Canaria, Spain

A R T I C L E I N F O

Article history:
Received 26 July 2005
Accepted 27 May 2011
Available online 17 June 2011

Keywords:
Real-time visual tracking
Heuristic tracking algorithm
Template matching
Template updating
Template tracking
Heuristic search
Target motion
Kullback–Leibler

A B S T R A C T

Many vision problems require fast and accurate tracking of objects in dynamic scenes. These problems can be formulated as exploration problems and thus can be expressed as a search into a state space based approach. However, these problems are hard to solve because they involve search through a space of transformations corresponding to all the possible motion and deformation. In this paper, we propose a heuristic algorithm through the space of transformations for computing target 2D motion. Three features are combined in order to compute efficient motion: (1) a quality of function match based on a holistic similarity measurement, (2) Kullback–Leibler measure as heuristic to guide the search process and (3) incorporation of target dynamics into the search process for computing the most promising search alternatives. Once 2D motion has been calculated, the result value of the quality of function match computed is used with the purpose of verifying template updates. A template will be updated only when the target object has evolved to a transformed shape dissimilar with respect to the actual shape. Also, a short-term memory subsystem is included with the purpose of recovering previous views of the target object. The paper includes experimental evaluations with video streams that illustrate the efficiency and suitability for real-time vision based tasks in unrestricted environments.

1. Introduction

Template tracking is a basic task in visual systems whose main goal is focused on detection and tracking a mobile object of interest in a dynamic vision context given one or several explicit templates that represent the target object. If an active vision approach is considered, it is also desirable that the tracking process keeps the object of interest centered in the image, moving the sensor adequately [1,2].

At present, there are still obstacles in achieving all-purpose and robust tracker approaches. Four main issues must be addressed in order to carry out an effective template tracking approach:

(1) *Real-time performance*. Real-time template tracking is a critical task in many computer vision applications such as vision based interface tasks [3], visual surveillance [35], traffic control [36], navigation tasks for autonomous robots [37], gesture based human–computer-interaction [38], perceptual intelligence applications [4], virtual and augmented reality systems [39] or applications from the "looking and people" domain [5]. Moreover, in real-time applications not all system resources can be allocated for tracking processes because other high-level tasks such as trajectory interpretation and reasoning can be demanded. Therefore, it is desirable to adjust the requirements of the computational cost of a tracker approach to be as low as possible to make feasible real-time performance over general purpose hardware.

(2) *Initialisation*. Many template based tracking approaches are focused on the use of a manual initialisation. Some approaches often assume that the template which represents the target object is correctly aligned in the first frame [6]. Other approaches select the reference templates by a hand-drawn prototype template, i.e., an ellipse outline for faces [7,8] or they are extracted from a set of examples such as level appearance [9] or outlines [10,11]. Moreover, the *condensation* algorithm [10] also requires training using the object moving over an uncluttered background to learn the motion model parameters before it can be applied to the real scene. However, these selection processes restrict its use in many practical embedded applications. Therefore, quick and transparent initializations without user participation are required.

(3) *Matching*. Template matching is the process in which a reference template $T(k)$ is searched for in an input image $I(k)$ to determine its location and occurrence. Over the last decade, different approaches based on searching the space of

☆ This work has been supported by the Spanish Government and Canary Islands Autonomous Government under the Projects TIN2004-07087 and PI2003/165.

* Corresponding author. Fax: +34 922473377.
E-mail addresses: enielsen@ull.es (E. Sánchez-Nielsen), mhernandez@iusiani.ulpgc.es (M. Hernández-Tej).

1047-3203/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jvcir.2011.05.005

transformations using a measurement similarity have been proposed for template based matching. Some of them explicitly establish point correspondences between two shapes and subsequently find a transformation that aligns these shapes [12,13]. The iteration of these two steps involves the use of algorithms such as iterated closest points (ICP) [14,15] or shape context matching [13]. However, these methods require a good initial alignment in order to converge, particularly whether the image contains a cluttered background. Other approaches are based on searching the space of transformations using Hausdorff matching [16], which are based on an exhaustive search that works by subdividing the space of transformations in order to find the transformation that matches the template position into the current image. Also, similar techniques have been used for tracking selected targets in natural scenes [30] and for person tracking using an autonomous robot [31]. However, no heuristic functions and no target dynamics have been combined in the search process. This situation leads to an increase of the computational costs in the tracking process.

(4) *Updating*. The underlying assumption behind several template tracking approaches is that the appearance of the object remains the same through the entire video [17–19]. This assumption is generally reasonable for a certain period of time and a naive solution to this problem is updating the template every frame [30–32] or every n frames [33] with a new template extracted from the current image. However, small errors can be introduced in the location of the template each time the template is updated and this situation establishes that the template gradually *drifts* away from the object [20]. Matthews et al. in [20] propose a solution to this problem. However, their solution approach only addresses the issue related to objects whose visibility does not change while they are being tracked.

In this paper, a template based solution for fast and accurate tracking of moving objects is proposed. The main contributions are focused on: (1) an A* search algorithm that uses the Kullback–Leibler measurement as heuristic to guide the search process for efficient matching of the target position, (2) dynamic update of the search space in each image, whose corresponding dimension is determined by the target dynamics, dramatically reducing the number of possible search alternatives, (3) updating templates only when the target object has evolved to a new shape change significantly dissimilar with respect to the current template in order to solve the drift problem and (4) representation of illustrative views of the target shape evolution through a short-term memory subsystem.

As a result, the first two contributions provide a fast algorithm to apply over a space of transformations for computing target 2D motion and the other two contributions provide robust tracking because accurate template updating can be performed. In addition to these contributions, the paper also contains a number of experimental evaluations and comparisons:

- A direct comparison of the performance of conventional search approaches [16] that work by subdividing transformations space and the proposed A* search approach that incorporates target dynamics and heuristic to guide the search process, demonstrating that A* search based approach is faster.
- An empirical comparison of updating templates using a continuous updating approach like that proposed in [30–32] and the updating template approach that is proposed in this paper, demonstrating that no updating templates in every frame and using a dynamic short-term memory subsystem, lead to a more robust tracking approach.
- An analysis of the time required for computing the proposed template matching and updating approach, illustrating that the time to track targets in video streams is lower than real-time requirements.

The structure of this paper is as follows: the problem formulation is illustrated in Section 2. In Section 3, the heuristic algorithm for computing target position is described. The updating reference template problem is detailed in Section 4. Experimental results are provided in Sections 5 and 6 concludes the paper.

2. Problem formulation

The template tracking problem of objects in 3D space from 2D images is formulated in terms of decomposing the transformations induced by moving objects between frames into two parts: (1) a 2D motion, corresponding to the change of the target position in the image space, which is referred to as the template position matching problem and (2) a 2D shape, corresponding to a different aspect of the object becoming visible or an actual change of shape in the object, which is referred to as the template updating problem.

For the sake of subsequent problem formulation, some definitions are first introduced:

Definition 1 (*Template*). Let $T(k) = \{t_1, \ldots, t_r\} \subseteq \mathbb{R}^2$ be a set of points that represent a template in step time k.

Definition 2 (*Image*). Let $I(k) = \{i_1, \ldots, i_s\} \subseteq \mathbb{R}^2$ be another set of points that denote an input image in step time k.

It is assumed that each new step time k corresponds to a new frame k of the video stream.

Definition 3 (*Set of transformations*). Let a translational transformation g be parameterized by the x displacement g_x and the y displacement g_y. That is, $g = (g_x, g_y)$.

Let a bounded set of translational transformations be a set of transformations $\mathbb{G} = [g_{xmin}, g_{xmax}] \times [g_{ymin}, g_{ymax}] \subseteq \mathbb{R}^2$ and let $g^c = (g_x^c, g_y^c)$ denote the transformation that corresponds to the center of $\mathbb{G}$. It is defined as:

$$g^c = \left(\left(\frac{1}{2}(g_{xmin} + g_{xmax})\right), \left(\frac{1}{2}(g_{ymin} + g_{ymax})\right)\right) \quad (1)$$

where ($xmin, xmax$) and ($ymin, ymax$) represent respectively the low and upper bounds of $\mathbb{G}$ in x and y dimension.

Definition 4 (*Bound error notion of quality of match*). Let a bounded error notion of quality of match $Q(g; T(k), I(k), \varepsilon)$ be a measurement for computing the degree of match between a template $T(k)$ and a current input image $I(k)$, where the dependence of Q on T, I and/or ε is omitted for sake of simplicity but without loss of generality. That is, the quality of match assigned to a transformation g is represented by the allowed error bound, ε, when template points are brought to image points using the transformation g. This quality of match function assigned to a transformation g is expressed as:

$$Q(g) = \sum_{t \in T} \max_{i \in I} \|g(t) - i\| < \varepsilon \quad (2)$$

where $\|\cdot\|$ denotes a measurement of distance and $g(t)$ represents the result of applying the transformation $g = (g_x, g_y)$ to every point in template $T(k)$.