Detection and Tracking through means of Consensus based Classification and Learning.

El Jeilany Sidi Mohamed

June 2015

Abstract

In this paper the author tries to address some of the issues of the currently available object tracking solutions; by proposing and describing a new robust and dynamic object tracking algorithm. The Proposed approach can handle changes in illumination, partial object occlusions, scale and in plane rotation variations and changes in object appearances; all while keeping false detections at low minimum.

Moreover, This approach algorithm can be applied to both, previously known and unknown objects; however this paper will focus on previously known and classified objects. The proposed method uses a three-steps approach. Consisting of a detection step followed by a consensus based matching and tracking step and finally a learning step.

1 Introduction

2 Related Work and Contributions

2.1 Object class detector

Object detection is the task of localization of objects of the same class in an input image. Object detection is a well studied subject in computer vision; and a multitude of solution for object detection exist. In the making of this paper the author has investigated three approaches for object detection[5],[3] and [8]. each one have it's own strong points and shortcomings.

The first approach was originally proposed by Voila *et al.* in [9] and later improved upon by Lienhart *et al* in [5].It uses Local Binary Features-based Cascade Classifier, its a widely used method mainly for face detection as its relatively inexpensive computational wise, the training time is also relatively short. LBF based classifiers usually yields good accuracy at the same time as a detection rate in the order of 98%. The major disadvantages of using (LBF and Haar) based classifiers is that they are not transformation invariant; while this method could be considered scale invariant to some extent its not rotation

invariant. Rotation invariance could be accomplished by training multiple classifiers but that would also decrease the accuracy and increase false detections and the computational cost.

[3] was introduced by Dalal *et al.* The technique uses grids of Histograms of Oriented Gradient descriptors, and a linear SVM classifier for object detection. This method is widely used for human and car detection. (HOG) features provide good accuracy sometimes better than (LBF and Haar) features. Like the previous approach this approach too does not provide rotation invariance.

[8] is a probabilistic method that uses Boosted random Ferns densely computed over local HOG space of binary features, in a two step approach. This method was introduced by Villamizar *et al*, It's two steps consist in an object pose estimator and an object classifier; which makes it rotation invariant when it comes to in plane rotation. This method is computationally inexpensive but yields lower detection rates that the two previous ones. Despite this, it's it's rotational invariance gives it an edge over the two.

Depending on the specific use case some Algorithms may be better suited than others. While an object detection method is not vital to Our proposed approach. However it's still essential as it provides an extra level of accuracy for the learning step and it helps with the tracking. Thus the following will assume that an appropriate object detection method is available.

2.2 Feature Detection, Description and Matching

Feature detection is a low-level image processing operation that allows the matching of local structures between images. Thus providing a sparse set of corresponding locations in different images. A multitude of Feature detection and description methods is available: Scale Invariant Feature Transform SIFT[6], Speed-up Robust Feature SURF[2], and more recently Oriented FAST and Rotated BRIEF ORB [7], Binary Robust Invariant Scalable Keypoints BRISK[4] and Accelerated AKAZE[1] to name a few.

3 Approach

given a stream of images Im_0 to Im_N our approach consist in a first time at characterizing all the appearances of a specific class (ei Car, People, boat...), using an object detection algorithm, and selecting a specific object O from that class; then at each frame recovering the pose of that specific object O by characterizing its { center: c ,scale: σ ,and in plane rotation: θ}. This section will describe the approach in details. By utilizing a static object detector combined with a dynamic object model, this approach aims at addressing the issues of changes in object appearances as well as false detections.

An algorithmic description of the inner-workings of the the proposed approach is detailed in Algorithm 1.

Algorithm 1 DTCCL

Input: $\{Im_i\}_{i=0}^{n}$
Output: $\{b\}_{i=0}^{n}$
{Initialization Stage}
$B \leftarrow object\_Detector(Im_0)$
$b_0 \leftarrow user\_Selection(B)$
$O \leftarrow kp\_DetectorAndDescriber(Im_0, b_0)$
$S \leftarrow springs\_Initialization(O, b_0)$
$W \leftarrow weight\_Initialization(O, b_0)$
$L \leftarrow linkage(S)$
$\Theta, D \leftarrow pairwise\_AngleAndDistance(L)$
$aKp \leftarrow O$
{Main Loop}
for all Im_i in $\{Im_i\}_{i=0}^{n}$ **do**
  $B \leftarrow object\_Detector(Im_i)$
  $K \leftarrow kp\_DetectorAndDescriber(Im_i)$
  $mKp \leftarrow kp\_Matcher(K, O)$
  $tKp \leftarrow kp\_Tracker(Im_i, Im_{i-1}, aKp)$
  $cKp \leftarrow fuse(mKp, tKp)$
  $L \leftarrow linkage(cKp)$
  $\theta, \sigma \leftarrow rotation\_AndScaleEstimation(L, \Theta, D)$
  $V \leftarrow voting(cKp, S, \theta, \sigma)$
  $\mu, inKp \leftarrow consensus(V)$
  $c \leftarrow verdict(\mu, \theta, \sigma, inKp, B)$
  if $c \neq \emptyset$ **then**
    $b_i \leftarrow bounding\_Box(\mu, \theta, \sigma, b_0)$
    $aKp, O, W, S \leftarrow learn\_NewKps(b_i,)$
  else
    $b_i \leftarrow \emptyset$
    $aKp \leftarrow \emptyset$
  end if
end for

3.1 Object Model

Object model O is a data structure of key-points and weights representing the object as observed so far.
$O = \{p_i, f_i, w_i\}_{i=0}^{k}$
Where $p_i \in \mathbb{R}^2$ represent the location of the key-point i expressed in term of local object coordinates. f_i represent the descriptor of the key-point i as computed by the chosen keypoint descriptor. And $w_i \in \{0,1\}^8$ its weight.

A separate data structure aKp is used to represent the active keypoints of O; aKp is used for optical flow tracking purposes. $aKp = \{r_i, id_i\}_{i=0}^{n}$ Where r_i is the position of the keypoint i in absolute image coordinates, and id_i is the index of i in O.

3.2 Initialization

We begin by initializing O. Using a keypoint-detection method, we detect and describe all keypoints in Im_0 that are inside the bounding box b_0 selected by the user after the object class detection step. Since most keypoint-detection methods give the kepoints positions in term of absolute image coordinates, we need to express O in term of local object coordinates. For optimisation purposes and to avoid repeating the same computation at each iteration of the main loop. One could compute two data-structures D and θ representing the pairwise distances and angles between the object keypoints.

3.3 Main Loop

In order to recover the pose of O in each $Im_i, i > 1$ we proceed as follow: We start by detecting all the

runing the object class detector on Im_i and storing the output bounding boxes in a datastructure B.
$B = \{cb_j\}_{j=0}^{k}$

3.4 Implementation Details

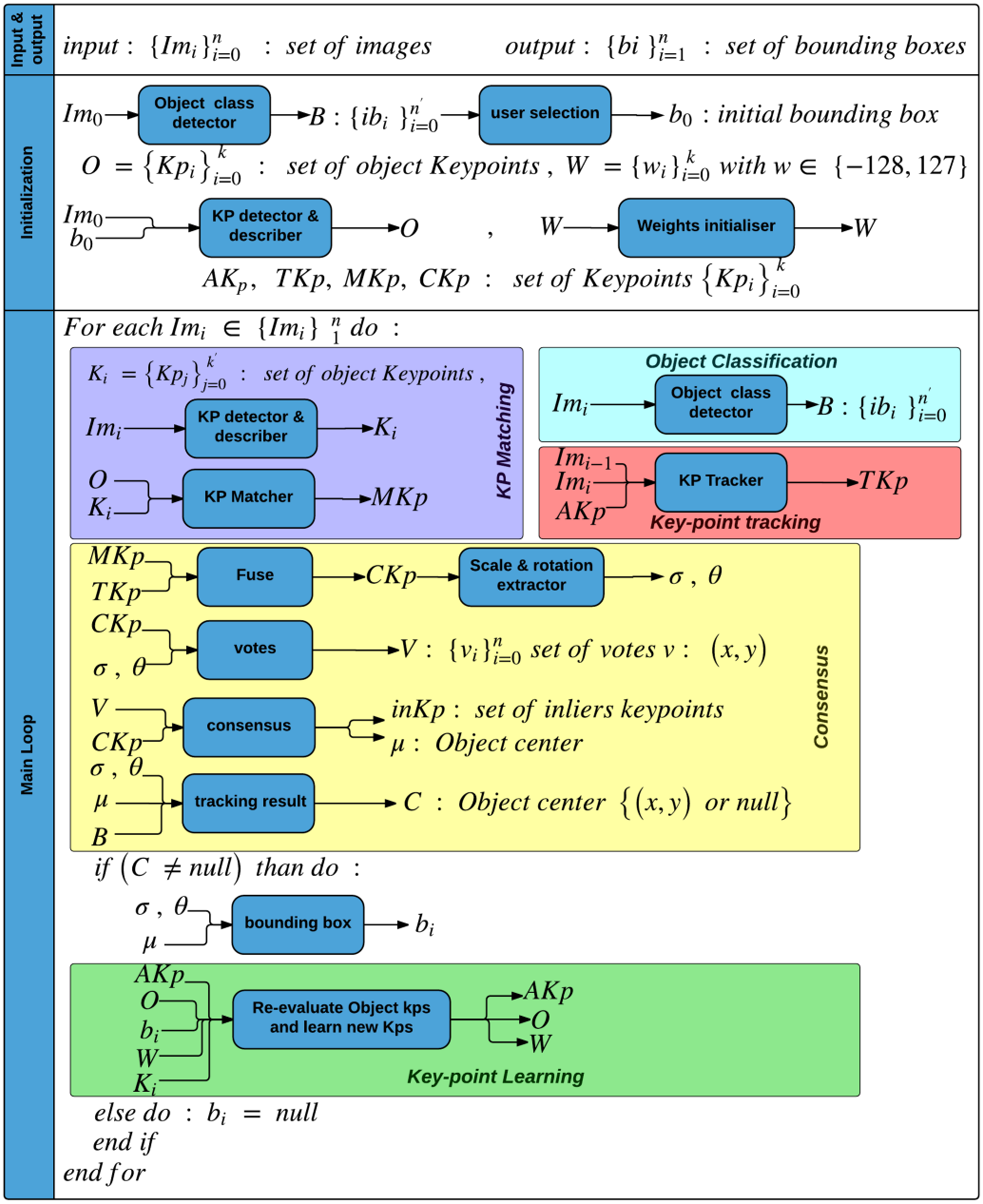

Figure 1: Approach diagram

[h]

References

[1] Pablo F Alcantarilla and TrueVision Solutions. Fast explicit diffusion for accelerated features in nonlinear scale spaces. *IEEE Trans. Patt. Anal. Mach. Intell*, 34(7):1281–1298, 2011.

[2] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *Computer vision–ECCV 2006*, pages 404–417. Springer, 2006.

[3] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.

[4] Stefan Leutenegger, Margarita Chli, and Roland Y Siegwart. Brisk: Binary robust invariant scalable keypoints. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2548–2555. IEEE, 2011.

[5] Rainer Lienhart, Alexander Kuranov, and Vadim Pisarevsky. Empirical analysis of detection cascades of boosted classifiers for rapid object detection. In *Pattern Recognition*, pages 297–304. Springer, 2003.

[6] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[7] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. Orb: an efficient alternative to sift or surf. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2564–2571. IEEE, 2011.

[8] Michael Villamizar, Francesc Moreno-Noguer, Juan Andrade-Cetto, and Alberto Sanfeliu. Efficient rotation invariant object detection using boosted random ferns. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1038–1045. IEEE, 2010.

[9] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–511. IEEE, 2001.