Development of the Tracking Database

Nina Wolleb

Introduction

Object tracking can have a number of practical applications. From surveillance of public areas to industrial production environments or human-computer interaction like an assistance robot where objects need to be tracked and manipulated. Whatever the application for the tracker is there are some difficulties appearing that make objects hard to track. These are sudden object motion, changing appearance pattern of object or scene, nonrigid object structures and occlusion regarding object or scene (Yilmaz et. al., 2006). A lot of tracking algorithms were proposed with the goal to overcome as many of these difficulties as possible. [examples nnbmic, esm bmic] Every research group creates and uses a own set of videos to test their tracking algorithm. This diversity of videos gives an impression on how good and how precise an algorithm works but does not lead to a direct comparability of different trackers. To overcome this we propose the Object Tracking Database containing videos showing different tasks recorded under different but consistent conditions and results of three different trackers (nearest neighbour tracker, SSD/Baker+Matthews inverse compositional tracker and efficient second order minimization tracker) running on these videos.

Related work

It is not that there are no video databases from other research groups. There is for example the INRIA Person Dataset (Dalal and Triggs, 2005) which was created for research work on detection of upright people and contains videos and images. Another extensive data collection is the Honda/UCSD Video Database (Lee et. al., 2005) which provides material for face tracking and recognition evaluation. The Metaio Benchmarking Dataset for Template Tracking Algorithms (Lieberknecht et. al., 2009) also allows a good comparison between trackers but has only eight datasets available. Also they show more abstract objects which are not very close to real life applications of tracking. These are only a fraction of examples from the existing ones out there. The main work found is restricted on face or pedestrian tracking and was created because it was needed for validation of research and published afterwards. Consequently their focus is not on a structured database with a broad range of videos that is convenient to use. Also we did not come over a database containing natural daily live manipulation tasks as proposed in our work.

Material and methods

All videos were recorded using a GRAS-20S4C-C firewire camera equipped with a Kowa LM6NCM F1.2/6mm lens. As a software programs and commands through ROS (Robot Operating System) and Python were used to record a .bag file video. The camera settings were implemented through Coriander 2.0.1 and varied between the diffuse and normal light condition recording as found in table 1. All videos were recorded at 30 frames per second and a resolution of 600x800 in YUV colour space. The camera stood on a fixed position on a tabletop 90 cm from the edge of the table.

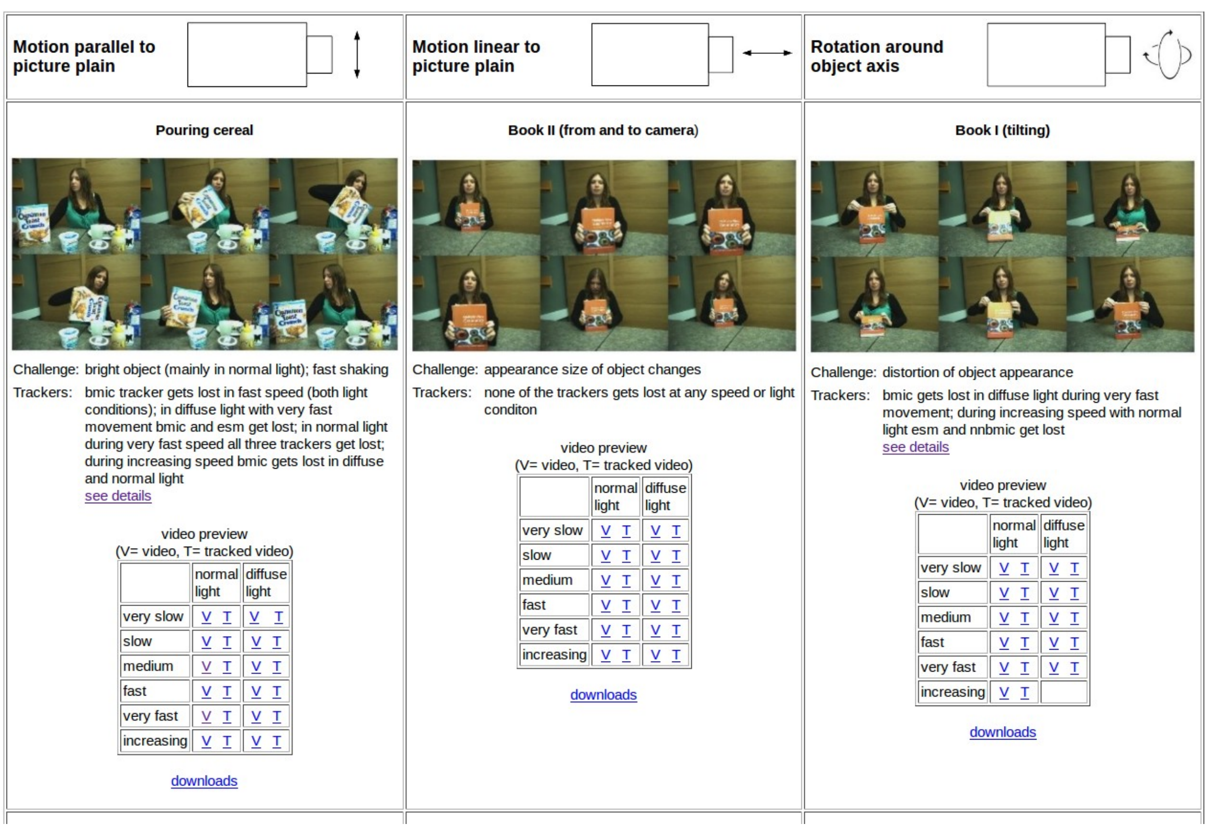
Table 1: Camera settings for diffuse and normal light condition:

	diffuse light	normal light
Exposure (in IL)	0.73	1.06
Gain (in dB)	6.02	0
Shutter (in s)	0.04	0.07
White Balance	Blue/U927 Red/V493	Blue/U757 Red/V490
Saturation (in %)	95.22	119.04

The starting point for finding different tasks was pouring cereal into a bowl which was first used to verify the nearest neighbour tracking algorithm [Dick et. al., 2012]. From there experimenting with different objects and different movements gained some experience in what would make sense and what would not and finally led to the following proposed tasks and variations.

Tracking complex tasks (e. g. tracking an envelope while putting a letter inside it) is very interesting because it is close to real world applications but for determining when and why a tracker fails complex tasks might not be the right choice. The reason for this is that if the tracker fails you can barely say why. It could be because of changing object appearance, a abrupt motion or the general speed. For this reason we split our videos into complex and oriented movement tasks.

Oriented tasks



Motion parallel to picture plain

Pouring cereal

Challenge: bright object (mainly in normal light); fast shaking

Trackers: bmic tracker gets lost in fast speed (both light conditions); in diffuse light with very fast movement bmic and esm get lost; in normal light during very fast speed all three trackers get lost; during increasing speed bmic gets lost in diffuse and normal light

see details

video preview
(V= video, T= tracked video)

	normal light	diffuse light
very slow	V T	V T
slow	V T	V T
medium	V T	V T
fast	V T	V T
very fast	V T	V T
increasing	V T	V T

downloads

Motion linear to picture plain

Book II (from and to camera)

Challenge: appearance size of object changes

Trackers: none of the trackers gets lost at any speed or light conditon

video preview
(V= video, T= tracked video)

	normal light	diffuse light
very slow	V T	V T
slow	V T	V T
medium	V T	V T
fast	V T	V T
very fast	V T	V T
increasing	V T	V T

downloads

Rotation around object axis

Book I (tilting)

Challenge: distortion of object appearance

Trackers: bmic gets lost in diffuse light during very fast movement; during increasing speed with normal light esm and nnbmic get lost

see details

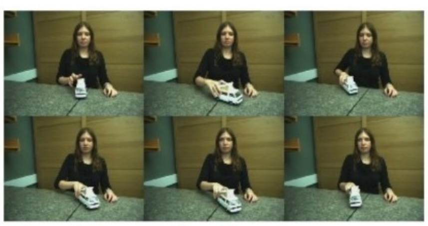
video preview
(V= video, T= tracked video)

	normal light	diffuse light
very slow	V T	V T
slow	V T	V T
medium	V T	V T
fast	V T	V T
very fast	V T	V T
increasing	V T	

downloads

Picture 1: Section of oriented tasks table.
The oriented movement tasks (picture 1) include different objects moved in different oriented ways to the camera which led to the subdivision of how the movement to the camera is. The categories here are parallel to picture plain, linear to picture plain and rotation. The book was chosen because of its highly textured surface. Movement from an to the camera (linear to picture plain) as well as rotation show well how the tracker reacts to a change in size and distortion. To see how the tracker handles occlusion in one task the book is inserted into a magazine file (parallel to picture plain). Tracking the chosen mug with its cylindrical shape and the mainly white and brown colour is challenging by itself. Moving parallel to picture plain, rotating and tilting challenges the tracker even more. Movement of the cereal box and the juice box with respect to the camera (parallel to picture plain) does not vary but the different textures as well as the rapid shaking of the cereal box while pouring give great differences to tracking. To add more opportunities to challenge the tracker all tasks were recorded in five different speeds as well as a video in which the speed is increasing during the task as tracking is sensitive to fast and abrupt motion of the object and can this way be tried in different speeds. These oriented, categorized tasks in different speeds lead to a high comparability between different trackers and give a good chance to detect why the tracker failed on a particular video. The consistent objects also contribute to comparability as the rigidity of the object does not change during different tasks.

Complex tasks

Bus	Highlighting	Letter
		
Challenge: little texture, change of appearance size	Challenge: nonrigid object	Challenge: little texture, distortion of object appearance
Trackers: in normal light none of the trackers fail; in diffuse light only esm survives; note that the movements are very different because of the complex motions carried out see details	Trackers: in normal light none of the trackers survives; in diffuse light all trackers survive; note that the there is a much bigger occlusion during the normal light video at the end see details	Trackers: in normal light none of the trackers fails; in diffuse light only nnbmic fails see details
video preview (V= video, T= tracked video) normal light: V T; diffuse light: V T	video preview (V= video, T= tracked video) normal light: V T; diffuse light: V T	video preview (V= video, T= tracked video) normal light: V T; diffuse light: V T
downloads	downloads	downloads

Picture 2: Complex task table.

Supplementary to the oriented movement tasks complex movement tasks (picture 2) were recorded which are a good way to test a tracker on a more sophisticated task that can be considered as assembled from various structured tasks. It needs to be mention that every complex movement task is available only in one speed because it is hard to reproduce such complex movements in varied speeds without making it untraceable. For the complex

movement videos no subdivision seemed necessary as there are only a few videos in total and subcategories are hard to find regarding the complexity of the tasks. One complex movement task is a miniature VW bus driving on a table top. The bus task is a challenge to the tracker as the bus is primarily white with little texture and through motions in different directions the appearance changes quickly. Another challenge for the tracker is putting a letter into a, even so it is cardboard enhanced, relatively nonrigid envelope. The highlighting task where things are tagged with a highlighter on newspaper contains sophisticated change of occlusion. Another video also involving a newspaper can be found. Here the newspaper is picked up, unfolded, read, folded up and placed on the tabletop again. The highly nonrigid newspaper is a big challenge to trackers. This variety of complex tasks should give a good impression on how good a tracker works.
When tracking the tasks sometimes the tracker likely failed because of reflections and glare of the object and not because of for example speed. For this reason videos are available in different light conditions. One half of the set of videos were recorded in normal office light which someone may find more commonly in real life. Therefore the picture is bright. Glare and reflections can be found on the objects. If videos without these difficult light conditions are needed to underline the difference in tracking the videos in diffuse light are available. These videos were recorded under a screen eliminating direct light from above. The disadvantage of the screen is less light. To compensate this darkness gain was added which leads to a noisy and still darkish video but without bigger reflections on the object.
All this above mentioned available videos of tasks and different distinctions lead to a fair number of videos to choose from if a tracker shall be tested.

To organize this fair number (almost 100) of videos and make them easily assessable a website was created. The videos are presented in tables. The oriented task table has headlines with the main categories of object movement in respect to the camera including a pictogram for visualization. Each task is in the table is named an a image with thumbnails can be found to see what the task is about at a first glance. Mainly the videos were presented in a table for each speed headlined with main categories describing the movement in respect to the camera as mentioned before. To guarantee visualization on first sight there are videos for in browser replay available. Another feature that can show what to expect from the video of a certain task is the rating of every task in the categories geometry, texture and size. The rating was created applying the experience gained from further trying of the tracker and from visible impressions of the video. As downloadable output files pooled in a .zip file we chose to make the video in form of images (.jpg) and as a video with the two mentioned trackers available. In addition a directly playable video can be found on the website to get a first impression on what is available so people do not have to download things they do not need in the end. Text in note form gives information about what the challenge of the particular task is and which tracker(s) fail during it. Clicking on “see details” leads to a .txt file with a table containing the frame number in which a tracker failed. In the following table Videos with or without trackers can be previewed. The quality of these is lower than the quality of the .jpg images that can be found under the link “downloads”. Besides the images a .txt file giving the coordinates of the tracking corners (tracking data) for all three trackers can be found.

When looking at how good the three trackers (nearest neighbour from Dick et. al., SSD/Baker+Matthews inverse compositional tracker, efficient second order minimization tracker from Benhimane et. al.) work it strikes that that they are not likely to fail at the lower speeds. For making these speeds more challenging a smaller tracking area could be chosen. In the videos we created it is always the big tracking area covering most of the object was chosen.

Sources

Yilmaz, A., Javed, O., and Shah, M. 2006. Object tracking: A survey. ACM Computing Surveys 38, 4, Article 13

Dalal, N., Triggs, B., 2005. Histograms of Oriented Gradients for Human Detection. International Conference on Computer Vision & Pattern Recognistion 2, page 886-893

Lee, K., Ho, J., Yang, M., Kriegman, D., 2005. Visual Tracking and Recognition Using Probabilistic Appearance Manifolds. Computer Vision and Image Understanding 99, page 303-331

Lieberknecht, S., Benhimane, S., Meier, P., Navab, N. 2009. A Dataset and Evaluation Methodology for Template-based Tracking Algorithms. ISMAR 2009.

T. Dick, C. Perez, A. Shademan, M. Jagersand. A supervised Learning Approach to Registration-Based Visual Tracking. 2012

S. Baker and I. Matthews. Equivalence and efficiency of image alignment algorithms. Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, 2001

S. Benhimane, E. Malis. Real-time image-based tracking of planes using Efficient Second-order Minimization. Intelligent Robots and Systems, 2004. (IROS 2004) Proceedings of 2004 IEEE/RSJ International Conference on