COOPERATIVE OBJECT TRACKING USING TWO MODALITIES

Ondřej Šiler
PhD Degree Programme (1), FIT BUT
E-mail: siler@fit.vutbr.cz

Supervised by: Renaud Péteri, Přemysl Kršek
E-mail: renaud.peteri@univ-lr.fr, krsek@fit.vutbr.cz

ABSTRACT

Methods of particle filtering based on color distribution can be used to track non-rigid objects moving in videos. They are robust in case of noise or partial occlusions. However, this method is sensitive to changes in the lightening conditions of the scene. This change does not affect all the different wavelengths of the electromagnetic spectrum: the thermal infrared range is insensitive to variation in lightening conditions. A method for automatic cooperation between the color and the infrared modalities is proposed in this work. Used in a cooperative way, both an infrared and a color camera allow object tracking under difficult conditions.

keywords: *object tracking, particle filter, cooperative tracking*

1 INTRODUCTION

Video tracking is the process of locating a moving object in space and time using a camera. It is widely used in computer vision applications. Nowadays surveillance cameras are installed in many security-sensitive areas such as railway stations, parking blocks, airports or banks to improve the safety. One of the most succesfull method for tracking non-rigid object is particle filter. It is indeed efficient for multi-modal and non-linear tracking, and also in case of clutter or partial occlusions.

The tracking method used for a color video is based on color distributions of the target. In case of lightening changes, this method has difficulties and loses the target's track. However, lightening modifications do not affect the image in IR spectrum so this modality could be use to track the target temporally. A question arises – why not to use only an IR camera. IR cameras have in general a lower spatial resolution than color cameras, moreover, they are much more expensive. The lower spatial resolution does not prevent tracking as far as the object is sufficiently distinguishable in the image. Mostly the goal is not only to pursue target, but also its subsequent recognition. IR images are not suitable for this purpose.

2 PARTICLE FILTERING

Particle filtering [1, 2] is very successful for non-linear and non-Gaussian estimation problems and is reliable in cases of clutter and during occlusions. It is a technique for implementing a recursive Bayesian filter. The key idea is to represent a required posterior density function by a set of random samples with associated weights and to compute estimates based on these samples and weights. As the number of samples becomes very large, this characterisation becomes an equivalent representation of the usual functional description of the posterior probability density function (i. e. pdf) and the particle filter approaches the optimal Bayesian estimate.

To apply the particle filter, we have to define the state space, the dynamical and observation models. State space represents a set of all information characterizing the tracked object in time t. In our case, we have defined the state space as an upright elliptic region. The ellipse is determined by the lenghts of its half axes and its position. The dynamical model describes the evolution of the process. It relies on positions and velocities estimated in the previous frames. Velocity is updated according to the difference of mean target position in the last frame and the frame before last. The new length of half axes of the ellipse is either the same as previous, or $\pm 10\%$ choosen randomly with normal distribution.

We want to apply the particle filter into two domains (visible and IR), thus we need two different observation models. Color and temperature histograms are used as target models as they achieve robustness against non-rigidity, rotation and partial occlusion. We have chosen the observation model similar to [3]. The distributions are determined inside an upright elliptic region. To increase the reliability of the distribution when boundary pixels belong to the background or get occluded, smaller weights are assigned to the pixels that are further away from the region center. Thus, we increase the reliability of the distribution when these boundary pixels belong to the background or get occluded.

In a tracking approach , the estimated state is updated at each time step by incorporating the new observations. Therefore, we need a similarity measure. A popular measure between two distributions is the Bhattacharyya distance. For more details, the reader could refer to [4].

3 PROPOSED METHOD FOR COOPERATIVE TRACKING USING VISUAL AND IR INFORMATION

Two video records are used to follow the motion of an object. One of the records is considered as the first modality, in most cases color video, where we want to track a target as long as possible and the other one, secondary, is complementary. We suppose that the conditions are not convenient for tracking in the first modality, but that it is possible to do so in the second one. Whenever the target gets lost in primary source, it starts to be tracked in the secondary one. The target is tracked there until it is possible to find it back in the primary video again.

Video registration is a necessary pre-step in order to be able to integrate the data obtained from different cameras. Spatial registration is the process of transforming the different sets of data into one coordinate system. Because of the relative proximity of two cameras, affine transformation is reasonable for the registration precission we need for tracking. Temporal registration contemporizes corresponding events. Video sequences are arranged in order to have the same frequency. They are synchronized by finding same event distinguishable in both videos.

The goal of the proposed method is to continuously track the target despite difficult conditions appearing in one of the modality (like change of lightening, or tracking behind a glass which is

opaque in certain IR wavelength range). Suppose we want to track an object in a color video: in case of problems, we then switch to the infrared sequence, where tracking is possible. We use the Bhattacharyya distance of color distributions [4] as a measure of similarity between candidates and reference model. When the target disappears, the color distribution changes and the Bhattacharyya distance between the reference model and estimated position histogram increases. If the distance exceeds a certain threshold, we consider the target to be lost. We then begin to track it in IR video. Starting position is set to previous correctly estimated position in color video that is, thanks to video registration, at the same coordinate system as IR video. Then we operate tracking in IR video while tracking in the visible is periodically tested. We try to track in color video every x^{th} frame and, to be sure, we keep tracking in IR video for y frames as well. If the Bhattacharyya distance is less than the threshold, we stop tracking in IR and continue only in visible video. The diagram of this method is shown in Fig. 1.
The same principle stands also when infrared video is the first modality and color video the second. With this approach, we don't have to process both videos simultaneously. We switch between modalities when needed and continue processing of one video sequence.

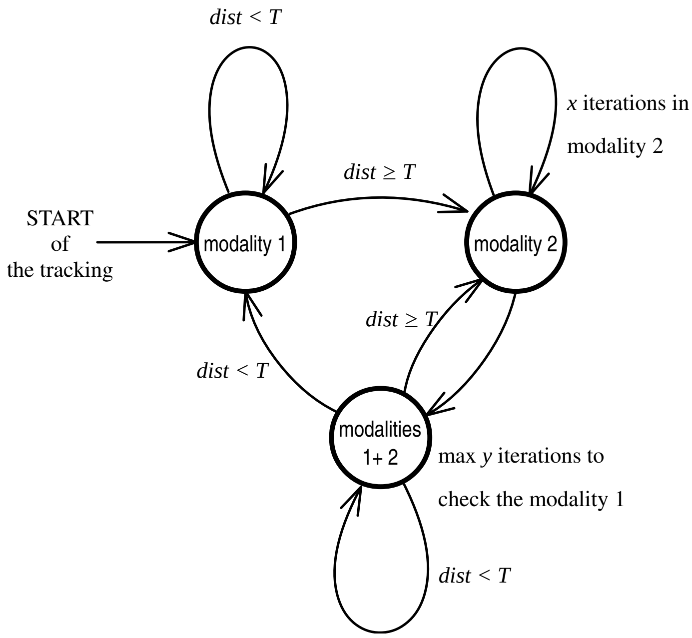

Figure 1: State diagram of the cooperative tracking method: tracking starts in the *modality 1*. After every processed frame, the Bhattacharyya distance *dist* is evaluated. If it is greater than the threshold T (T is relative to the modality 1) then tracking continues for x frames in the *modality 2*. Afterwards, both videos are processed (*modalities 1 + 2*) for a maximum of y frames. During this trial period, Bhattacharyya distance T is again computed in modality 1. If it does not exceed the threshold T, then the tracking comes back to *modality 1*, otherwise it returns back to *modality 2*.

4 WRITING RECOGNITION IN CHANGING LIGHTENING CONDITIONS

In the experiment we are trying to follow the trace of a heated lighter. A person writes a word with the lighter in the air in front of him. Lights are switched off and on during writing. If it is tracked only in the visible video, it looses the trace in the dark. When combined with IR video, we are able to continuously follow the lighter and recognize the word “infrared” (Fig. 2). The Bhattacharyya distance graph is shown in Fig. 3.

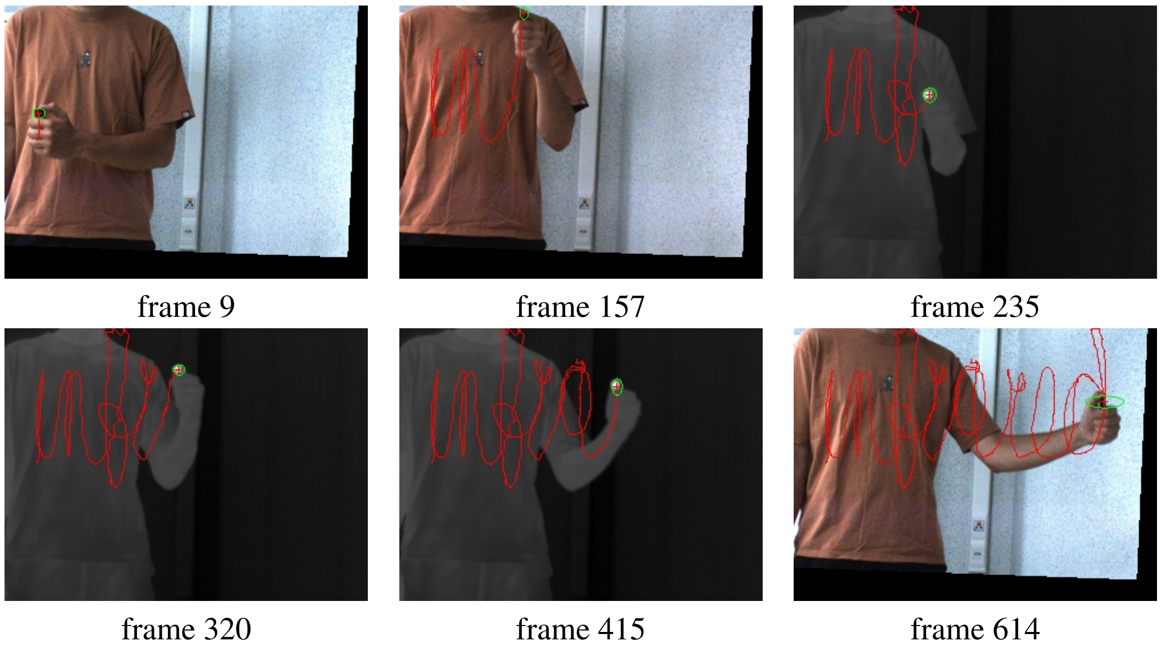

Figure 2: Writing tracking in difficult lightening conditions. While lights are on, we can track first two letters (up to frame 157). Tracking is preserved with the help of IR video in the dark period (frames 235 to 415). IR output is dispayed as grayscale images where brighter colors represent higher temperatures. Frames from color video have black borders at the bottom and the right side. They were rotated and cropped at the registration step but the FOV of color camera did not cover the whole scene taken by IR camera. Thus the missing parts are black.

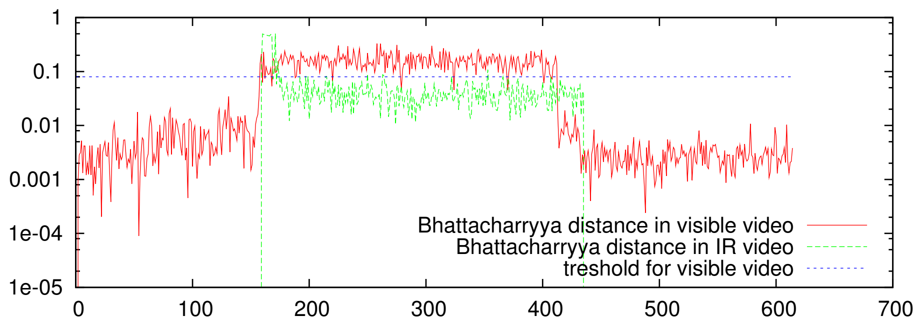

Figure 3: Bhattacharyya distance graph while cooperative tracking is performed. As long as lights are on the Bhattacharyya distance of estimated state in visible video is less than the threshold (till frame 158). From frame 159 the distance is above the threshold. In this period tracking is ensured in IR video. Few frames in visible video are tested periodically. It appears as decreases in Bhattacharyya distance in visible. During the last trial period (from frame 423 to 436) the Bhattacharyya distance of estimated target is always below the threshold (therefore tracking algorithm switches back to visible video).

5 CONCLUSION

This article describes a method for tracking a moving object with the help of two cameras: one color and one thermal infrared. A particle filter is used as a tracking algorithm in both spectra. It searches for identical or very similar color or temperature distribution of the target model. In our case, the target model remains the same during all the sequence. The model may be updated continuously, but the switching between modalities can cause the lost of the model validity. During the duration of the second modality, the target model is not updated. In case the target changes in the duration of the second modality, the color distributions differ. I.e. the model color (IR) distribution, that remains constant during tracking in the second modality, and the target color (IR) distribution after the return to the first modality are different. For this reason, the continuous model update was not used. However there is a possibility it can refine the tracking method efficiency.

One of the most important part in the algorithm is to decide when is the right time to switch to the other modality. In the project, it is decided according to a Bhattacharyya distance threshold which is set empirically. How to determine automatically that the target is lost is subject for further researches. Similar task relates to particle filter parameters that have to be correctly adjusted to obtain good results. Now they are set empirically as well, as it is ofen the case in particle filter methods.

The cooperative tracking method could work not only with color and infrared cameras. For example, two color cameras at different locations, with different points of view, can be used as well. Then an object can be tracked despite large occlusions. In this case, video registration should be improved, because affine transformation can not work this complex problem out.

REFERENCES

[1] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, *A tutorial on particle filters for on-line non-linear/non-gaussian bayesian tracking*, 2002.

[2] M. Isard *Visual motion analysis by probabilistic propagation of conditional density*, 1998.

[3] K. Nummiaro, E. Koller-Meier, and L. V. Gool, *Color features for tracking non-rigid objects*, Special Issue on Visual Surveillance, ACTA Automatica Sinica (Chinese Journal of Automation), (2003).

[4] N. A. Thacker, F. J. Aherne, and P. Rockett, *The bhattacharyya metric as an absolute similarity measure for frequency coded data.*, Kybernetika, 34 (1997), pp. 363–368.