Czech Technical University in Prague

Doctoral Thesis Statement

Czech Technical University in Prague
Faculty of Electrical Engineering
Department of Cybernetics

Tracking with Context

Lukáš Cerman

PhD Programme: Electrical Engineering and Information Technology
Branch of Study: Artificial Intelligence and Biocybernetics

Doctoral thesis statement for obtaining the academic title of “Doctor”, abbreviated to “Ph.D.”

Prague, February 2013

The doctoral thesis was produced in a full-time manner during Ph.D. study at the Center for Machine Perception of the Department of Cybernetics of the Faculty of Electrical Engineering of the CTU in Prague.

Candidate: Ing. Lukáš Cerman
Department of Cybernetics
Faculty of Electrical Engineering of the CTU in Prague

Thesis Advisor: Prof. Ing. Václav Hlaváč CSc.
Department of Cybernetics
Faculty of Electrical Engineering of the CTU in Prague

Opponents: ....
....
....

The doctoral thesis statement was distributed on ....

The defence of the doctoral thesis will be held on

.... at .... a.m./p.m.

before the Board for the Defence of the Doctoral Thesis in the branch of study Artificial Intelligence and Biocybernetics in the meeting room No. .... of the Faculty of Electrical Engineering of the CTU in Prague.

Those interested may get acquainted with the doctoral thesis concerned at the Dean Office of the Faculty of Electrical Engineering of the CTU in Prague, at the Department for Science and Research, Technická 2, 166 27 Prague 6.

Prof. Ing. Vladimír Mařík DrSc.

Chairman of the Board for the Defence of the Doctoral Thesis
in the branch of study Artificial Intelligence and Biocybernetics
Faculty of Electrical Engineering of the CTU in Prague
Department of Cybernetics
Karlovo náměstí 13, 121 35 Prague 2

Contents

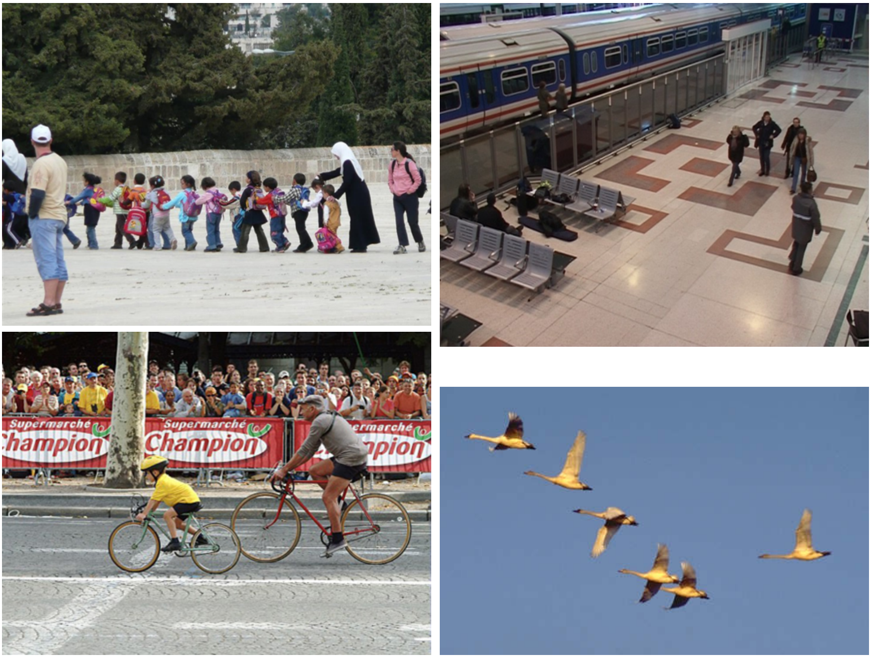

Figure 1: Objects in a context.

1 Introduction

Visual tracking has been massively studied in the last two decades with many application areas like surveillance, driving assistance or movie production industry. Our observation is that in the majority of the traditional approaches only the object itself and/or its background are modeled. However objects rarely move independently of its neighborhood and are embedded into a context that is often ignored. There may exist other parts of the scene that are not part of the object itself but exhibit, temporal or permanent, motion correlation to object. See Figure 1 showing examples of the situations where a strong link exists between the object and the other parts of the observed scene. Tracking the each individual object – a child, a pedestrian, a cyclist or a bird – independently in those situations may be more difficult than tracking them as a group.

Discovering the link between the object and the other parts of the scene – let us call them companions – and tracking them together can improve

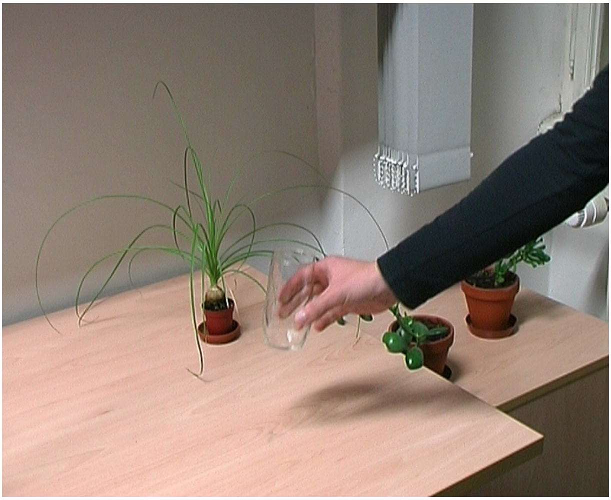

Figure 2: Where is the glass?

the robustness of the tracker in the situations that would be difficult for the traditional tracker. Those may include cases, in which the object's appearance changes quickly and significantly or even the situations where the object is not directly observable due to the occlusion or it is behind the border of the image. Tracking some objects without a companion can be very difficult, e.g, a glass held in hand as shown in Figure 2.

The environment is not stable usually in the real applications, e.g., illumination or view may change, the object appearance changes with its motion, there may be clutter and occlusions. A tracker that would not react to those changes can easily fail or may be applicable only in very constrained environments. To broaden its applicability many trackers include some adaptation mechanisms reflecting the changing reality. The actual presence of a companion in the scene may affect the adaptation of the tracker that is not aware of it. The unrecognized companion may randomly merge with the object model causing the model to not describe the object anymore. It may also be assigned to the background distracting its model. For this reason, looking for a companion and explicitly modeling it may be not only good in supporting the tracker in difficult situations but also for obtaining the better models that match the real situation more closely.

The visual tracking is usually approached using the appearance features in various forms like contours, colors, texture, etc., either in the generative form: by creating the models that describe the appearance of the object and/or its background; or in the discriminative form: by training a classifier that uses the image-based features to distinguish the object from the background. However, there is another feature, which is rarely used in tracking, the motion. It is independent of the appearance and naturally complement it. Adding the motion-based features to the subject model can extend the range of situations the tracker can handle. It may then be able to distinguish the object from the background even if they look similarly by observing the difference in their motion and, with the same models, it may be able to discriminate the still object from the still background by their differing appearance.

Goals of the Thesis

The aim is to design a robust visual tracker. This can be achieved by pushing the state of the art in several aspects, that have been mainly overlooked in the past. Informally, the goals of the thesis can be summarized as looking for the answers to the following questions: *"What to model?"*, *"Which features to choose to build the models?"*, and *"How to combine the features of different modalities?"*.

1. *"What to model?"* The first question is mainly related to studying of the object's context, which has not been used much in the state of the art approaches. Usually, only the object itself and/or its background is modeled. It should be explored what are the benefits of modeling and using the object's context in tracking. The object's context cannot be known in advance. Usually, especially in the single object tracking, the only data known is the object bounding box at the beginning of the video sequence. A method that would allow to learn the context on-line is requested.

2. *"Which features to choose to build the models?"* The majority of the state of the art approaches uses only subject's appearance in the tracker reasoning. Our second goal is to improve the subject's models with the use of a motion features in addition to the widely used appearance features.

3. *"How to combine the features of different modalities?"* This is also related to the second question – the features coming from the different domains, such as appearance and motion, should be combined together to allow the inference. The ways of this combination should be studied.

2 Contributions

The contributions of the thesis can be summarized to three points:

1. We have proposed a novel image-based tracker called the Sputnik Tracker. The tracker is able to discover which regions of the observed image exhibit the same motion as the target object. This information is then used to stabilize the tracking in the situations that would be difficult for a common tracker, e.g., strong appearance changes of the target object or its occlusions.

2. We have proposed a histogram-like model called the hierarchical histogram, which can be used to model RGB color of subject's in many computer vision tasks. Unlike the traditional 3-D histogram it can be estimated from the limited training data without the need to reduce the data precision.

3. We have formulated tracking as a semi-supervised learning and labeling task, in which the tracked feature points are labeled to the three classes – an object, a background and a companion, which represents the object's context. The use of Markov random fields allows a simultaneous integration of the appearance, motion and shape features. Similarly to the Sputnik Tracker, the use of the context, represented here by the companion class, allows the tracker to estimate the object position even if it is not directly observable or undergoes a strong appearance changes. The addition of the motion features allows the tracker to distinguish the object from the background if they look similarly.

3 State of the Art

Foreground and Background Trackers

The approaches to tracking can be divided to several groups by the source of information that is used to estimate the position of the tracked object. One group is formed by the approaches that are based on the background segmentation, often called background subtraction, which is a process of dividing the image area to segments recognized as a background and segments recognized as a foreground (objects). Some authors [14, 22, 36, 26, 3, 23, 38, 27, 28] use directly the results of the motion segmentation to perform the tracking. Others use it just to constrain the other tracking methods [4] or perform tracking on the segmented image instead of on the original image data [19, 12, 13]. By definition, the background subtraction methods handle only the background model while completely ignoring the appearance of the object.

Another group of approaches, called appearance template trackers [2, 6, 15] is based solely on modeling the object while ignoring the background. In those approaches, a generative representation of an object – a template – is created before the tracking starts. The template can be either fixed or it can be dynamically updated during the tracking process to reflect variations in the object's appearance. The inference of the object position is then performed by searching for the best matching position of the template in the image. This search is usually performed as a minimization of some cost function.

Our point is that a robust tracker should take care of both. This could help in situations, in which, for instance, the objects changes its appearance so strongly that it no longer fits the foreground model. The traditional template tracker would fail in such a situation. The tracker that uses the background model in addition to the foreground one may still be able to estimate the object position by recognizing the background region occluded by the object. The necessary condition is that the new appearance of the object differs from the background, if not, the additional features, like motion, would be needed.

From this point of view, our approach can be related to the layer-based methods [34, 35, 29, 30], which try to explain the whole image, not just the object or background area, and the classifier-based approaches, which discriminate between the background and foreground [33, 7, 1, 16], or the discriminative template trackers [5, 20].

Use of Context in Tracking

The role of the context has been mainly studied in relation to the object detection [31, 21, 9, 11]. For instance, it has been shown that a shadow that a vehicle casts on the road is a good predictor of its position [32].

The ways of using the context in the object tracking area are less explored. It has been addressed, for instance, by Grabner et al. [18]. In their work the context information is represented as a set of image points, called supporters, that are repeatedly observed in the image. Each of the points votes for the object position. A strength of the vote is proportional to the correlation between the translation of the point and the translation of the object centroid. The set of supporters is learned and updated on-line during the tracking process. The authors demonstrated that the use of supporters allowed to track the object behind an occlusion.

Dinh et al. [10] further developed this idea by adding a set of distractors. Distractors are regions which have a similar appearance as the target and consistently co-occur with the high confidence score. The tracker must keep tracking these distractors to avoid interchanging them for the object by a mistake, which can, for instance, happen when the object is occluded in presence of the distractor. The supporters are used in a similar way as in [18].

Yang et al. [37] proposed to explore a set of auxiliary objects that have a strong motion correlation and a co-occurrence with the target in a short term. Also, they need to be tracked easily. The auxiliary objects are represented by the image regions that are obtained using a color segmentation and are described by a color histogram. The Meanshift is applied to track them.

4 Sputnik Tracker

We describe a method similar to the layer-based approaches [34, 35, 29, 30] adapted for the purpose of a single object tracking. It uses only two layers, the first one is attached to the object and the second one represents the background. Other objects, if present, are not modelled explicitly. They may become a part of the foreground layer and represent the object's contex or become a part of the background outlier process. Such approach can be also viewed as a generalized background subtraction combined with an appearance template tracker.

The image-based representation of both foreground and background, inherited from the layer-based approaches, contrasts with the statistical representations used by classifiers [17] or the discriminative template trackers [5, 20], which do not model the spatial structure of the layers. The inner structure of each layer can be useful source of information for localizing the layer.

The key observation is that the tracked object is often accompanied by some other objects that move coherently with the object. By discovering those objects, called companions, and including them to the extended foreground layer, we obtain a more robust tracker. The connection of the object to the companion may be temporary, e.g., a glass can be picked up by a hand and dragged from the table, or it may be permanent, e.g., a head of a man always moves together with his torso, see Figure 3 for examples. As the core contribution, we show how the context represented by

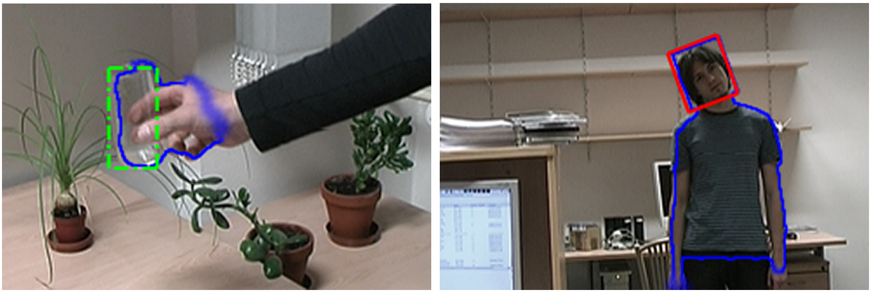

Figure 3: Objects with a companion: foreground includes not just the main object, e.g., a glass (on the left) or a head (on the right), but also other image regions, such as a hand or a body.

the companion improves the tracking and expands the set of situations, in which a successful tracking is possible without the need to model object's dynamic. We show that with the help of companion and the knowledge of the background it is possible to track the object that is not directly visible or the object that is rapidly changing its appearance. This would be very difficult for the conventional trackers that look only for the object itself. It is the companion what defines the position of the foreground layer in such situations.

The Sputnik Tracker is able to track a single part of a bigger object. However, unlike the methods based on the pictorial structures [13, 24, 25], it does not need the prior knowledge of the object structure, i.e., the number of the moving parts and their connections. The remaining parts of the object would be automatically assigned to the foreground layer as a companion provided that their motion is correlated with the motion of the object.

Results

To show the performance of the Sputnik Tracker, two sequences that would be challenging for a common tracker are presented. In all following figures, the red rectangle is used to illustrate a successful object detection, a green rectangle corresponds to the recognized occlusion or the change of object appearance. The blue line shows the contour of the foreground layer including the estimated companion. The thickness of the line is proportional to the uncertainty in the layer segmentation. The complete sequences as well as some other sequences can be watched on-line at the thesis companion

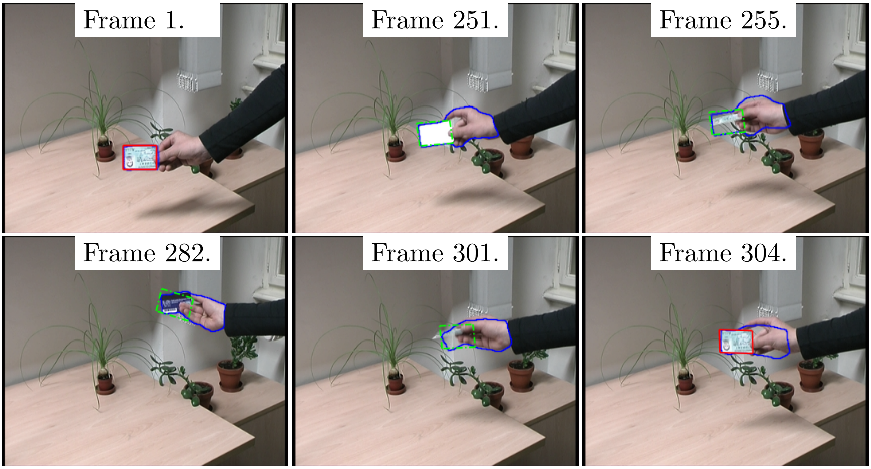

Figure 4: Tracking card carried by the hand. The strong reflection in frame 251 or flipping the card later does not cause the Sputnik Tracker to fail.

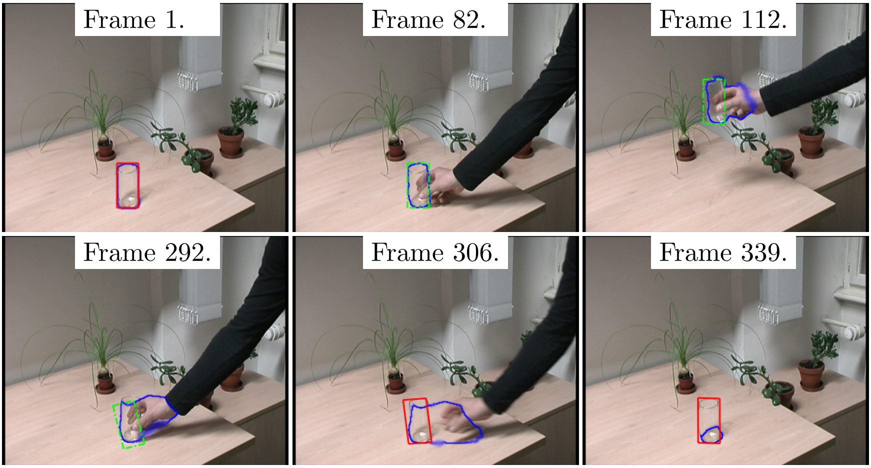

Figure 5: Tracking a glass after being picked by a hand and put back later. The glass moves with the hand which is recognized as companion which enables tracking of the transparent object.

web page *http://doiop.com/tracking-with-context*[1]or downloaded as video files from the same place.

The first sequence shows the tracking of an ID card, see Figure 4 for several frames selected from the sequence. After initialization with the region belonging to the card, the Sputnik Tracker learns that the card is accompanied by the hand. This prevents it from failing in the frame 251 where the card reflects strong light source and its image is oversaturated. Any tracker that looks only for the object itself would have a very hard time at this moment. Similarly, the knowledge of the companion helps to keep a successful tracking even when the card is flipped in the frame 255. The appearance on the backside differs from the frontside. The tracker recognizes this change and reports an occlusion. However, the rough position of the card is still maintained with respect to the companion. When the card is flipped back it is redetected in the frame 304.

Figure 5 shows tracking of a glass being picked by a hand in the frame 82. At this point, the tracker reports an occlusion that is caused by the fingers and the hand is becoming a companion. This allows the tracking of the glass while it is being carried around the view. Because of the glass transparency, its appearance vary with the background, making it a very difficult object to track without a companion. The glass is dropped back to the table in the frame 292 and when the hand moves away it is recognized back in the frame 306.

5 Tracking as a Semi-Supervised Learning and Labeling Problem

The idea of tracking with a companion is further developed by reformulating the problem. Instead of the image based model we propose to build a 3-D (space-time) graph from the independently tracked feature points. Each point in the graph gets assigned to an object, background or companion class so that the energy of the Markov random field defined on the graph is minimized.

As already mentioned in the introduction, the individual subjects, i.e., object, companion and background, can be distinguished by the difference in their appearance and/or the difference in their motion in the video sequence. The appearance and the motion features naturally complement each other. The Markov random field (MRF) is an elegant and conceptually simple framework allowing to integrate the features of different modalities, such as the appearance and the motion, in the model. It also allows to impose constraints on the subjects' shapes and the temporal stability of the subjects' regions. Estimates and updates of the involved models are possible on-line, which is an important property, since the context cannot be known in advance and the appearance of the object and/or the background can evolve during tracking.

We show how to formulate the tracking as a labeling problem using the MRF and describe the involved probabilistic models that enabled an algorithm based on the simple MRF framework to successfully track objects in the real sequences. We also demonstrate that the decision to use the MRF in the tracking algorithm does not make it computationally infeasible and that, with a careful selection of the involved probabilistic models and the implementation design, it can perform close to real-time on an ordinary PC that is common in the year 2013.

Results

Figure 6 shows tracking results of a mobile phone in 1359 frames long sequence, in which the phone gets occluded several times, and exhibits strong reflections on its shiny surface. The whole video with the results as well as the additional sequences can be seen on the thesis companion web page: *http://doiop.com/tracking-with-context*.[1]

In the figure, a green cross (×) depicts a point labeled as an object, red cross (×): background, yellow circle (○): companion, magenta circle (○): outlier moving with a foreground, and blue circle (○): other outliers. The textureless areas are not covered by feature points therefore appear void in the presented images. If the object gets occluded, its expected position is visualized by a green ellipse. There is no dynamics involved, the position is estimated using the motion of the companion only.

The proposed tracker is able to recognize the object in the whole sequence. It even survives the long full occlusion spanning frames 1144 to 1187. This improves our previous results [c], where the track was lost at this position due to the absence of the long-term appearance model, which helps to overcome such a situation. With the previous approach, the only way to overcome the long occlusions was to increase the size of the sliding window in the time domain. With the long-term model, we have succeeded tracking the object in the same sequence with only three frames long window as can be seen on the companion web page referenced above. Images presented here were obtained with the fifty frames long window.

[1] This is a shortened URL, if it does not work, you may try the original URL: *http://cmp.felk.cvut.cz/∼cermal1/tracking-with-context*.

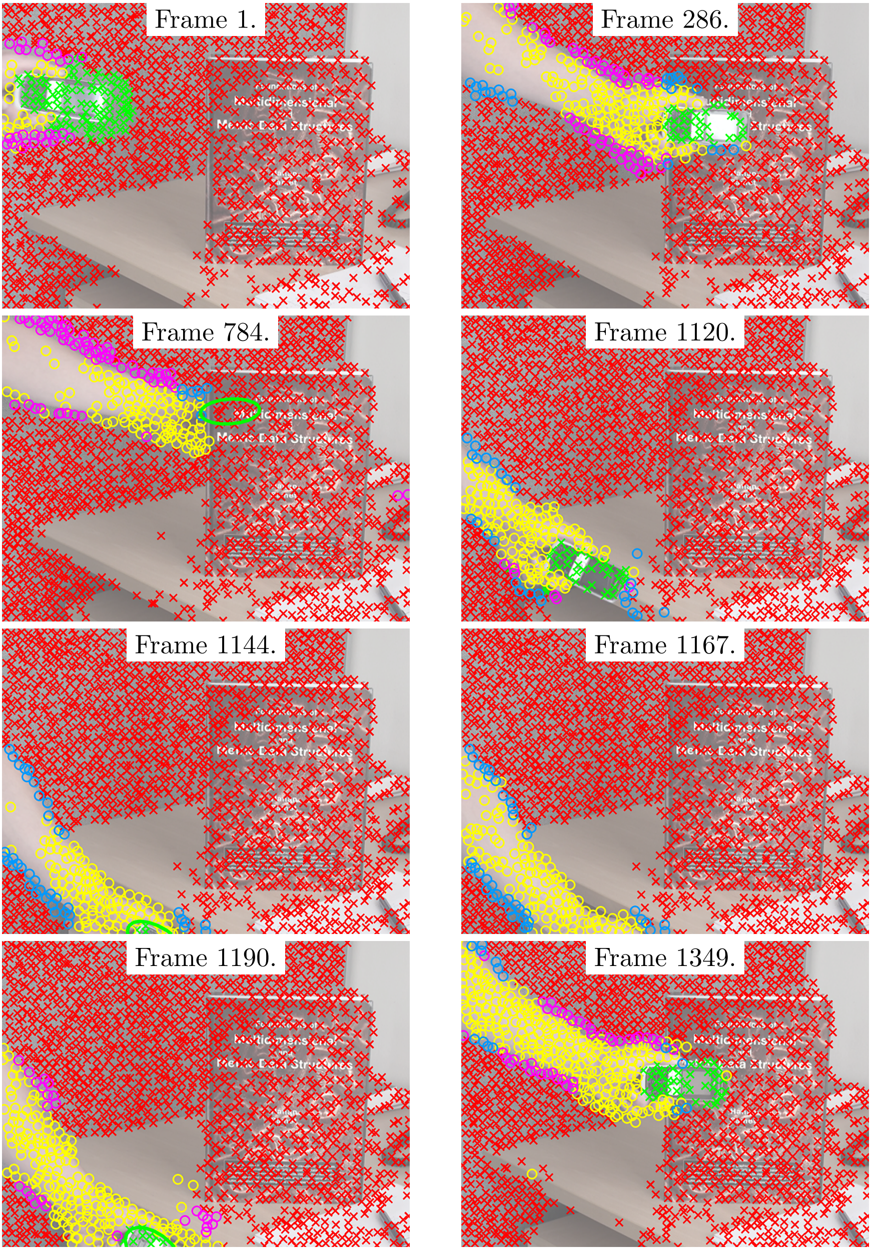

Figure 6: Sample sequence, for the description see the text.

6 Hierarchical Histogram Model

To model the appearance of subjects's in the later of the two proposed tracking algorithms a novel histogram like model has been proposed. It is called the hierarchical histogram model (HHM) and is able to represent the underlaying distribution with the similar quality as the frequently used Gaussian mixture model but it surpasses it in the speed of parameters estimation and the log-likelihood evaluation. Unlike traditional histograms, the HHM compromises on the *model size* vs. the *precision* dilemma by providing a fine probability density estimate in the areas largely supported by the data and a coarse estimate in the areas having sparse data support.

7 Conclusions

We have shown that the context of an object, which has been mainly overlooked by the state of the art approaches for a long time, plays the important role in the object tracking. The explicit modeling and identification of the context, which is represented by a companion in our work, can help the tracker to overcome difficult situations while ignoring an actual presence of the companion may even distract the tracker.

To demonstrate this, we have implemented the Sputnik Tracker, which is outlined in *Section 4*, and presented a successful tracking in several challenging video sequences. The tracker is based on a novel template tracker simultaneously evaluating foreground and background appearance cues. In addition, it learns on-line which image regions accompany the object and maintains an adaptive model of the companion appearance and its shape. This makes it robust to situations that would be distractive to trackers focusing only on the object alone. It includes situations in which the object is not directly observable.

A histogram-like model was suggested for the representation of multi-dimensional distributions such as RGB colors of subjects in tracking and segmentation tasks. Unlike the ordinary 3-D histogram it can be estimated from the limited amount of training data without the need to reduce the precision of the measured data. The proposed hierarchical histogram model is able to represent the underlaying distribution with the similar quality as the widely used Gaussian mixture model but it surpasses it in the speed of parameters estimate and log-likelihood evaluation. This makes it a suitable color model for the time critical tasks like the real-time tracking.

In the last part of the thesis, we have further developed the idea of tracking with a companion. In addition, we have studied a possible way of integrating the motion features together with the appearance features, which naturally complement each other. Including the motion features in the subjects' models extends the set of situations the tracker can handle to situations, in which the object looks similarly to the background. It also makes the tracker more robust to the variations in the subjects' appearance.

We have suggested how to formulate the tracking task as a semi-supervised learning and labeling problem, in which the independently tracked feature points are labeled to the three classes: the object, the background and the companion. The proposed approach is based on the Markov random fields, which allows an elegant integration of the appearance, motion and shape features in a single objective function. The algorithm requires a little user interaction – only the selection of the object points in the first video frame – to automatically recognize the object, background and companion points in the rest of the video sequence. This was demonstrated on multiple challenging sequences that include the partial and full occlusions, the appearance variations caused by the surface reflections or even a camouflage.

Future Work

The semi-supervised learning and labeling problem can be generalized for more classes, not just one object, its companion and the background. This might involve scenes with a background composed of the multiple differently moving regions or applications to the simultaneous tracking of multiple objects. It would be also interesting to introduce a completely unsupervised version that would not require any user interaction. The multiple object and background regions would be recognized automatically based on their distinct appearance and/or motion. This may by achieved by following the ideas from [8].

It would be also interesting to see this algorithm operating in a real-time as a whole. This task involves reimplementing all components that have not yet been written in C++ into C++. As a part of this reimplementation, some components might need to be simplified to allow smooth running on the contemporary computer architecture. The real-time performance would broaden the applicability of the proposed algorithm and would open new questions that might be interesting for the future research.

References

[1] S. Avidan. Support vector tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(8):1064–1072, October 2004.

[2] R. V. Babu, P. Pérez, and P. Bouthemy. Robust tracking with motion estimation and local kernel-based color modeling. *Image and Vision Computing*, 25(8):1205–1216, 2007.

[3] J. Black and T. Ellis. Multi camera image tracking. *Image and Vision Computing*, 14(11):1256–1267, November 2006.

[4] M. Bray, P. Kohli, and P. H. S. Torr. PoseCut: Simultaneous segmentation and 3D pose estimation of humans using dynamic graph-cuts. In *Proceedings of the European Conference on Computer Vision*, volume 3952 of *LNCS*, pages 642–655. Springer, 2006.

[5] R. Collins, Y. Liu, and M. Leordeanu. Online selection of discriminative tracking features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1631–1643, 2005.

[6] D. Comaniciu, V. Ramesh, and P. Meer. Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(5):564–575, 2003.

[7] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, volume 1, pages 886–893, Washington, DC, USA, 2005. IEEE Computer Society.

[8] A. Delong, A. Osokin, H. N. Isack, and Y. Boykov. Fast approximate energy minimization with label costs. *International Journal of Computer Vision*, 96:1–27, 2012.

[9] C. Desai, D. Ramanan, and C. Fowlkes. Discriminative models for multi-class object layout. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 229–236, Kyoto, Japan, October 2009.

[10] T. B. Dinh, N. Vo, and G. Medioni. Context tracker: Exploring supporters and distracters in unconstrained environments. In *Proceedings of the International Conference on Vision and Pattern Recognition*, pages 1177 –1184. IEEE Computer Society, 2011.

[11] S. K. Divvala, D. Hoiem, J. H. Hays, A. A. Efros, and M. Hebert. An empirical study of context in object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1271–1278, Miami Beach, Florida, June 2009. IEEE.

[12] L. Fajt. Pictorial structural models, learning and recognition in image sequences. Master's thesis, Czech Technical University, Center for Machine Perception, 2007.

[13] P. F. Felzenschwalb and D. P. Huttenlocher. Pictorial structures for object recognition. *International Journal of Computer Vision*, 61(1):55–79, 2005.

[14] L. M. Fuentes and S. A. Velastin. People tracking in surveillance applications. *Image and Vision Computing*, 14(11):1165–1171, November 2006.

[15] B. Georgescu, D. Comaniciu, T. X. Han, and X. S. Zhou. Multi-model component-based tracking using robust information fusion. In *Proceedings of 2nd Workshop on Statistical Methods in Video Processing*, pages 61–70. Springer, 2004.

[16] H. Grabner and H. Bischof. On-line boosting and vision. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, volume 1, pages 260–267, 2006.

[17] H. Grabner, M. Grabner, and H. Bischof. Real-time tracking via online boosting. In *Proceedings of the British Machine Vision Conference*, volume 1, pages 47–56, 2006.

[18] H. Grabner, J. Matas, L. Van Gool, and P. Cattin. Tracking the invisible: Learning where the object might be. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pages 1285–1292, San Francisco, USA, June 2010. IEEE Computer Society.

[19] F. Korč. 2D model-based tracking of humans in a single view sequence. Master's thesis, Czech Technical University, Center for Machine Perception, 2006.

[20] M. Kristan, J. Pers, M. Perse, and S. Kovačič. Closed-world tracking of multiple interacting targets for indoor-sports applications. *Computer Vision and Image Understanding*, 113(5):598–611, 2009.

[21] H. Kruppa. *Object Detection using Scale-specific Boosted Parts and a Bayesian Combiner*. PhD thesis, Swiss Federal Institute of Technology, Zurich, 2004.

[22] L. Marcenaro, L. Marchesotti, and C. S. Regazzoni. Self-organizing shape description for tracking and classifying multiple interacting objects. *Image and Vision Computing*, 14(11):1179–1191, November 2006.

[23] C. Motamed. Motion detection and tracking using belief indicators for an automatic visual-surveillance system. *Image and Vision Computing*, 14(11):1192–1201, November 2006.

[24] D. Ramanan. Learning to parse images of articulated bodies. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems*, pages 1129–1136. MIT Press, 2006.

[25] D. Ramanan, D. A. Forsyth, and A. Zisserman. Tracking people by learning their appearance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(1):65–81, 2007.

[26] A. Senior, A. Hampapur, Y. Tian, L. Brown, S. Pankanti, and R. Bolle. Appearance models for occlusion handling. *Image and Vision Computing*, 14(11):1233–1243, November 2006.

[27] C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, volume 2, page 252, 1999.

[28] C. Stauffer and W. E. L. Grimson. Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):747–757, 2000.

[29] H. Tao, H. S. Sawhney, and R. Kumar. Dynamic layer representation with applications to tracking. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, volume 2, pages 134–141. IEEE Computer Society, 2000.

[30] H. Tao, H. S. Sawhney, and R. Kumar. Object tracking with Bayesian estimation of dynamic layer representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):75–89, 2002.

[31] A. Torralba. Contextual priming for object detection. *International Journal of Computer Vision*, 53(2):169–191, 2003.

[32] C. Tzomakas and W. Von Seelen. Vehicle detection in traffic scenes using shadows. Technical report, Institut fur Nueroinformatik, Ruhr Universitat, 1998.

[33] P. Viola and M. Jones. Robust real-time object detection. In *Proceedings of the International Workshop on Statistical and Computational Theories of Vision*, Vancouver, Canada, July 2002.

[34] J. Y. A. Wang and E. H. Adelson. Layered representation for motion analysis. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pages 361–366. IEEE Computer Society, 1993.

[35] Y. Weiss and E. H. Adelson. A unified mixture framework for motion segmentation: Incorporating spatial coherence and estimating the number of models. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pages 321–326. IEEE Computer Society, 1996.

[36] M. Xu and T. Ellis. Augmented tracking with incomplete observation and probabilistic reasoning. *Image and Vision Computing*, 14(11):1202–1217, November 2006.

[37] M. Yang, Y. Wu, and G. Hua. Context-aware visual tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(7):1195–1209, July 2009.

[38] Q. Zhou and J. K. Aggarwal. Object tracking in an outdoor environment using fusion of features and cameras. *Image and Vision Computing*, 14(11):1244–1255, November 2006.

List of Candidate's Publications

Publications Related to the Thesis

Impacted journal articles

[a] Lukáš Cerman, Václav Hlaváč. Learning object's context helps the tracking formulated as a three-class semi-supervised learning and labeling problem. Submitted for consideration in *International Journal of Pattern Recognition and Artificial Intelligence*, February 2013. World Scientific Publishing Company. *IN REVIEW.* Authorship: 50-50.

Publications excerpted by WOS

[b] Lukáš Cerman, Václav Hlaváč and Jiří Matas. Sputnik Tracker: Looking for a companion improves robustness of the tracker. In *Proceedings of the Scandinavian Conference on Image Analysis*, pages 291–300, Oslo, Norway, June 2009. Springer-Verlag. Authorship: 34-33-33.

[c] Lukáš Cerman, Václav Hlaváč. Tracking with context as a semi-supervised learning and labeling problem. In *Proceedings of 21st International Conference on Pattern Recognition*, pages 2124–2127, Tsukuba, Japan, November 2012. IEEE Computer Society. Authorship: 50-50.

Other conference/tech-report publications

[d] Lukáš Cerman, Jiří Matas and Václav Hlaváč. Robust tracking: template matching meets background subtraction. In *Computer Vision Winter Workshop*, Moravske Toplice, Slovenija, February 2008. Slovenian Pattern Recognition Society. Authorship: 34-33-33.

[e] Lukáš Cerman, Karel Zimmermann and Tomáš Pajdla. Human detection and tracking with restart. *Research Report*, CTU–CMP–2010–24, Center for Machine Perception, Czech Technical University in Prague, December 2010. Authorship: 34-33-33.

[f] Lukáš Cerman, Václav Hlaváč. Hierarchical histogram for RGB color modeling. In *Proceedings of the Computer Vision Winter Workshop*,

pages 32–38, Hernstein, Austria, February 2013. Vienna University of Technology. Authorship: 50-50.

Another Publications

Other conference/tech-report publications

[g] Lukáš Cerman, Václav Hlaváč. Exposure time estimation for high dynamic range imaging with hand held camera. In *Proceedings of the Computer Vision Winter Workshop*, pages 76–81, Telč, Czech Republic, February 2006. Czech Society for Cybernetics and Informatics. Authorship: 50-50.

[h] Lukáš Cerman and Akihiro Sugimoto and Ikuko Shimizu. 3D shape registration with estimating illumination and photometric properties of a convex object. In *Proceedings of the Computer Vision Winter Workshop*, Graz, Austria, February 2007. Institute for Computer Graphics and Vision, Graz University of Technology. Authorship: 34-33-33.

Citations of Candidate's Work

[i] Helmut Grabner, Jiri Matas, Luc Van Gool, and Philippe Cattin. Tracking the invisible: learning where the object might be. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1285–1292, San Francisco, USA, June 2010. IEEE Computer Society.

[ii] Kee-Hyon Park, Dae-Geun Park, and Yeong-Ho Ha. High dynamic range image acquisition from multiple low dynamic range images based on estimation of scene dynamic range. *Journal of Imaging Science and Technology*, 53(2), 2009.

[iii] R. Ramirez Orozco, I Martin, C. Loscos, and P-P Vasquez. Full high-dynamic range images for dynamic scenes. In *Proceedings of the Conference on Optics, Photonics and Digital Technologies for Multimedia Applications II*, volume 8436, Brussels, April 2012. International Society for Optics and Photonics.

[iv] Rene Restrepo, Nestor Uribe-Patarroyo, and Tomas Belenguer. Im-

provement of the signal-to-noise ratio in interferometry using multi-frame high-dynamic-range and normalization algorithms. *Optics Communications*, 285(5):546–552, 2012.

[v] Zhongqian Sun, Hongxun Yao, Shengping Zhang, and Xin Sun. Robust visual tracking via context objects computing. In *Proceedings of the IEEE International Conference On Image Processing*, pages 509–512, Belgium, September 2011. IEEE Signal Processing Society.

[vi] Diego Thomas and Akihiro Sugimoto. Robustly registering range images using local distribution of albedo. *Computer Vision and Image Understanding*, 115(5):649–667, 2011.

[vii] Anna Tomaszewska and Radoslaw Mantiuk. Image Registration for Multi-exposure High Dynamic Range Image Acquisition. In *Processing of the International Conference in Central Europe on Computer Graphics*, pages 49–56, January 2007, Plzeň, Chech Republic. Union Agency Science Press.

[viii] Javier Vargas, Thomas Koninckx, Juan Antonio Quiroga, and Luc Van Gool. Three-dimensional measurement of microchips using structured light techniques. *Optical Engineering*, 47(5), 2008.

Resumé in Czech

Dizertační práce se zabývá studiem kontextu objektů v souvislosti se sledováním objektů ve videosekvenci. Kontext je v současných sledovacích metodách zřídkakdy využíván a pokud ano, tak většinou v souvislosti s detekcí objektů. V dizertační práci ukážeme, že kontext může být přínosem také pro sledovací algoritmy

Navrhneme dva sledovací algoritmy, které využívají kontext ve svém rozhodování. Oba algoritmy jsou schopné odhalit v obraze oblasti, které nejsou součástí objektu, nicméně jejich pohyb je s pohybem objektu silně svázán. Tyto oblasti, nazývané souputníky, jsou oba navržené algoritmy schopné odhalit v průběhu sledování a tuto znalost poté využít ke zlepšení kvality sledování v obtížných situacích, např. když není sledovaný objekt přímo viditelný nebo se jeho vzhled prudce mění.

První algoritmus, nazvaný Sputnik Tracker, je založený na obrazových šablonách. Algoritmus v průběhu sledování rozpoznává, které oblasti obrazu reprezentují souputníky, a zahrnuje je do šablony popředí, která je použita spolu s šablonou pozadí k robustnímu odhadu pozice objektu.

Druhý algoritmus využívá namísto obrazových šablon význačných bodů detekovaných v obraze, které jsou odsledovány nezávislým sledovacím algoritmem. Sledování je v tomto případě formulováno jako úloha učení a značkování s částečně označenými daty, ve které jsou význačné body rozděleny značkováním do třech tříd – objekt, pozadí a souputník. Sledovaný objekt je potom reprezentován shlukem bodů, jimž je přiřazena třída objekt. Formulace úlohy založená na Markovských náhodných polích umožňuje současné využití vzhledu i pohybu jako příznaků sloužících k modelování subjektů. Přidání pohybu do využívaných modelů umožňuje odlišit objekty od pozadí, i když vypadají podobně.

Model vzhledu použitý ve druhém algoritmu je založen na novém hierarchickém histogramu, který může být využit i v jiných úlohách počítačového vidění, jež vyžadují modelování barev. Výhodou navrženého modelu je oproti běžnému třírozměrnému histogramu možnost odhadu jeho parametrů z malého množství dat bez nutnosti snížit jejich přesnost.