Real-Time Object Tracking with Relevance Feedback

Ard Oerlemans, Joachim T. Rijsdam, Michael S. Lew
LIACS Media Lab, Leiden University, The Netherlands
{aoerlema, jrijsdam, mlew}@liacs.nl

ABSTRACT

Currently there are many systems available that use relevance feedback for text and image retrieval. This query by example method has been shown to optimize the search strategy whilst keeping a fast response time, two important factors when querying large image databases. The use of relevance feedback in real time video object tracking and identification however is essentially unexplored. This paper discusses the ongoing project and current results towards integrating interactive relevance feedback within the context of video object tracking. We discuss the important limitations in real time video object tracking and we design a next generation video object tracking system which exploits interactive relevance feedback towards addressing the primary limitations.

Categories and Subject Descriptors

I.4.8 [**Image processing and computer vision**]: Scene analysis – *tracking*.

General Terms

Algorithms, Performance

Keywords

Video analysis, motion detection, object tracking, relevance feedback

1. INTRODUCTION

Tracking moving objects in images from a stationary camera, such as people walking by or cars driving through a scene, has gained much attention in the recent years because it is a very useful technique for human-computer interaction, the next-generation games, and for surveillance applications.

We developed a video object server that can analyze live input streams from any video input device and which is able to output the locations, unique identifiers and pictorial information of the moving objects in the scene. The server is HTTP-based so that any web browser would be sufficient to display the output. Figure 1 illustrates the basic interface.

Object detection and identification however is a topic that has its own unique set of problems that still are not fully addressed. Multiple object tracking is complicated by the fact that objects can touch or interact with each other, can occlude another, even leave the scene and come back again. And the usual problems with single object tracking, like illumination changes or object deformation, still apply as well.

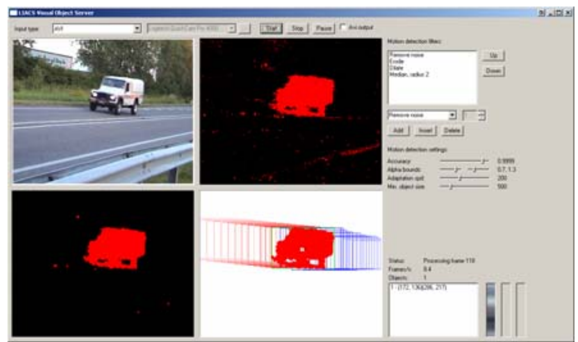

Figure 1. Liacs Visual Object Server interface.

To get improved object tracking results, we investigate methods to include user feedback for detecting moving regions. We have therefore extended our object tracking framework to include color-based and soon also texture-based relevance feedback functionality.

2. RELATED WORK

Extensive work has been done in the field of object tracking in video [1-10]. Notable scientific meetings include the Performance and Evaluation of Tracking and Surveillance (PETS), and Video Surveillance & Sensor Networks (VSSN). The articles and presentations from these conferences conclude that tracking multiple objects is a complicated process, not only due to object interactions and sensor restrictions like the Field of View or color balances, but also due to the challenges that single object tracking still poses, challenges like illumination and background changes, single object deformation, articulation and occlusion.

Many of the current systems achieve good accuracy by making special assumptions like restricting the system to just a few objects, or indoor scenes only. In fact, the conclusion of some of the advanced systems is simply that they are unsuited for outdoors [8].

Object tracking [3,6,8,10] depends on motion detection within a scene. There are a multitude of methods on how to detect this, ranging from standard background subtraction

CIVR'07, July 9-11, 2007, Amsterdam, The Netherlands.

methods [5] based on color or intensity, to sophisticated methods like Gaussians or optical flow [14].

For the object tracking itself, there are also a number of methods available. The choice of the right method depends on the application. Basic object tracking can be done using simple Nearest Neighbor matching of regions with motion. An improvement of this method would be motion prediction, for example with Kalman filters [11] or Bayesian networks [4].

Relevance feedback [12] is an interactive method that has been successfully introduced into text and image queries. It is an interactive query process where a computers internal representation of the object that a user is interested in is continually updated by providing the system with information about the relevancy of the query outcome.

In most algorithms in the research literature several important factors have been essentially neglected. First, the identity of the object over time is typically ignored. This includes both grouping subparts into one object and reporting misdetection rates for object parts when they become misclassified as new objects. Second, rigorous tests on different types of occlusion are also ignored in that there do not even exist test sets with ground truth for the occlusion problem.

For a more in-depth survey of multimedia information retrieval in general, see [14].

3. MOTION DETECTION

Horprasert, Harwood and Davis have developed a method [1] for robust motion detection using background subtraction based on a statistical model of the background. The main idea of their method is to decompose the difference between the current color information and the modeled background for every pixel into a chromaticity (color) component and a brightness component. Figure 2 illustrates this idea by plotting the mean RGB value, *E*, of a pixel and the current value, *I*. Given these two values, the brightness distortion α and chromaticity distortion *CD* can be calculated.

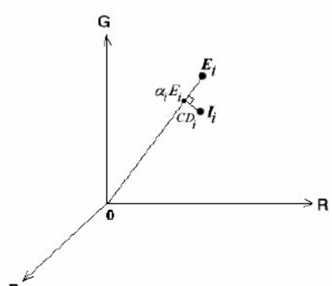

Figure 2. The brightness distortion α and chromaticity distortion *CD* between pixel values *E* and *I*.

For pixel *i*, given the current RGB values I_R, I_G, and I_B, the mean values μ_R, μ_G and μ_B, and the standard deviations σ_R, σ_G and σ_B, the formula to calculate the brightness distortion is:

$$\alpha_i = \min\left[\left(\frac{I_R(i)-\alpha_i\mu_R(i)}{\sigma_R(i)}\right)^2 + \left(\frac{I_G(i)-\alpha_i\mu_G(i)}{\sigma_G(i)}\right)^2 + \left(\frac{I_B(i)-\alpha_i\mu_B(i)}{\sigma_B(i)}\right)^2\right]$$

$$= \frac{\left(\dfrac{I_R(i)\mu_R(i)}{\sigma_R^2(i)} + \dfrac{I_G(i)\mu_G(i)}{\sigma_G^2(i)} + \dfrac{I_B(i)\mu_B(i)}{\sigma_B^2(i)}\right)}{\left(\left[\dfrac{\mu_R(i)}{\sigma_R(i)}\right]^2 + \left[\dfrac{\mu_G(i)}{\sigma_G(i)}\right]^2 + \left[\dfrac{\mu_B(i)}{\sigma_B(i)}\right]^2\right)}$$

The chromaticity distortion can then be calculated with:

$$CD_i = \sqrt{\left(\frac{I_R(i)-\alpha_i\mu_R(i)}{\sigma_R(i)}\right)^2 + \left(\frac{I_G(i)-\alpha_i\mu_G(i)}{\sigma_G(i)}\right)^2 + \left(\frac{I_B(i)-\alpha_i\mu_B(i)}{\sigma_B(i)}\right)^2}$$

With these two values and a threshold, determined by an automated threshold selection method as described in [1], each pixel can be classified as either being foreground or background.

One assumption that the authors of this motion detection method made, was that the lighting would stay roughly constant. In real world applications however, the light can change gradually. Thus, we implemented an adaptive version of their model to compensate for dynamic real world lighting. Small parts of the background model are continuously updated with new data, to compensate for these gradual changes in lighting. Another effect of this adaptation is that deposited objects can be added to the background model if they do not move for a given period of time.

We use bounding boxes as a method for object tracking. Objects are considered either simple or compound. Simple objects are objects that can be tracked by the straightforward version of the object tracker, in which every blob corresponds to no more than one object. In the case of object interactions, or overlapping objects, there is ambiguity as to which blob belongs to which object. We define compound objects as virtual objects which consist of two or more objects which are currently interacting in some fashion. The object tracker will track these objects just as simple objects, but will also track the actual simple objects which are involved.

A set of rules for tracking is applied to group simple objects into compound objects and vice versa. The details of this method can be found in our previous work [16].

3.1 Relevance feedback for motion detection

Our object tracking system continuously compares pixel values to the modeled background to decide whether a pixel should be considered as part of a moving object. The relevance feedback component can change this decision by learning from user input. An example is given in figure 3. In this case, the user indicates that objects that look like the selected part of the image should never be considered to be an object, even if the background model indicates that it should, which could happen in case of fast lighting changes.

Figure 3. User feedback to the oject segmentation algorithm: selecting a negative example.

The user can supply feedback while the tracking system is running. The selected positive and negative examples are immediately included in the decision process, so the effect on the object tracking results is instantly visible.

In figure 4 an overview is given of the object tracking system with relevance feedback module.

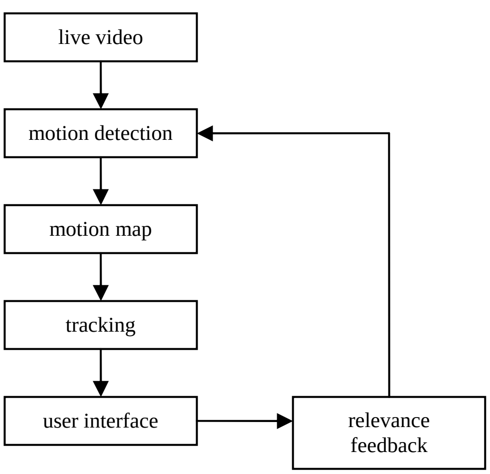

Figure 4. The relevance feedback module in the object tracking system.

3.2 User feedback with color features

Below is an example of using color as a relevance feedback feature. Figure 5 shows a frame from a sequence in which a person is leaving an object behind. In figure 6, the user selects a positive example for the object segmentation algorithm. In this case, objects with equal color will always remain marked as foreground and they will not be added to the background model, which is the normal behaviour for adaptive object segmentation algorithms. Figure 7 shows the object being classified as a foreground object based on the user feedback.

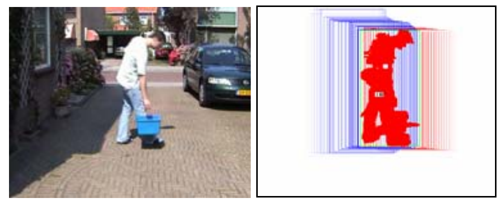

Figure 5. One frame from a sequence where someone leaves an object behind, together with the object tracking results.

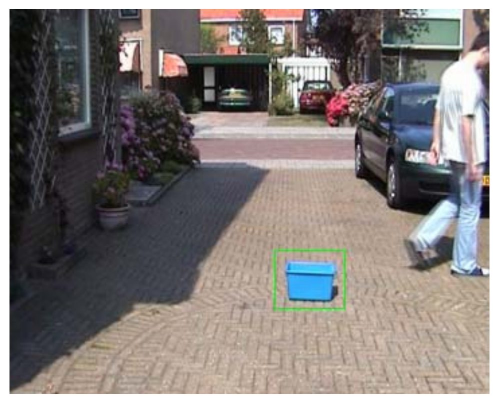

Figure 6. The user selects a positive example for the object segmentation algorithm.

Figure 7. The object segmentation using the positive example. The object will not be added to the background model and stays visible as a tracked object.

Negative examples using color features are useful for marking objects that are actually part of the background. In object segmentation with background modeling, many tracking errors are made because of quick changes in the real background that the background model cannot keep up with. In this case, supplying negative examples to the segmentation algorithm can help in deciding if a detected object should still be considered as background.

3.3 User feedback with texture features

The next step in our investigation into relevance feedback for object tracking, is using texture features. Although a preliminary version is already working, real results are not yet generated.

Texture features are somewhat more difficult to include in the object segmentation, because texture is not bound to single pixels, as color is. We have therefore assigned each pixel a texture value based on a small square region around it. A texture feature can then still be compared on a pixel-level basis. We hope to analyse the results of the texture based feedback in the very near future.

4. FURTHER RESEARCH

Our object tracking system shows that including relevance feedback in the object tracking process can be useful, although no quantitative results are available yet. We are continuing our research efforts to determine what other features could also be used. A shape feature would be a possible next step, but combining several features might be an interesting area to look into.

References

[1] T. Horprasert, D. Harwood, and L.S. Davis, "A statistical approach for real-time robust background subtraction and shadow Detection," IEEE Frame-Rate Applications Workshop, Kerkyra, Greece, 1999.

[2] B. Gloyer, H.K. Aghajan, K.Y. Siu, and T. Kailath, “Video-based freeway monitoring system using recursive vehicle tracking”, IS&T-SPIE Symposium on Electronic Imaging: Image and Video Processing, 1995.

[3] I. Haritaoglu, D. Harwood, and L. Davis, “W4: real-time surveillance of people and their activities”, IEEE Trans. on Pattern Analysis and Machine Intelligence, Volume 22 No. 8, Pages 809-830, 2000.

[4] B. Han, Y. Zhu, D. Comaniciu, L. Davis, “Kernel-based Bayesian filtering for object tracking”, Proc IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005.

[5] M. Cristani, M. Bicego, and V. Murino, “Integrated region- and pixel-based approach to background modeling”, Proc. of IEEE Workshop on Motion and Video Computing, Pages 3-8, 2002.

[6] A.Colombari, A. Fusiello, and V. Murino, “Segmentation and tracking of multiple objects in video sequences”, Proc. of the 5th International Workshop on Image Analysis for Multimedia Interactive Services, 2004.

[7] L. Li, W. Huang, I. Gu, and Q. Tian, "Foreground object detection from videos containing complex background," in ACM Multimedia, 2003.

[8] N.T. Siebel, and S. Maybank, “Real-time tracking of pedestrians and vehicles”, Performance Evaluation of Tracking and Surveillance PETS 2001 (with CVPR 2001), Kauai, Hawaii, Dec 9th, 2001. CD-ROM proceedings.

[9] Q. Xiong, and C. Jaynes, “Multi-resolution background modeling of dynamic scenes using weighted match filters”, Proceedings of the ACM 2nd international workshop on video surveillance & sensor networks, New York, NY, USA. Pages 88-96, 2004

[10] S.J. McKenna, S. Jabri, Z. Duric, and A. Rosenfeld, “Tracking groups of people”, Computer Vision and Image Understanding, Volume 80 No. 1, Pages 42-56, October 2000.

[11] N. Funk, “A study of the kalman filter applied to visual tracking”, University of Alberta, 2003

[12] Y. Rui, and T.S. Huang, “Relevance feedback techniques in image retrieval”, APR Principles of Visual Information Retrieval, Springer Verlag, Pages 219-258, 2001

[13] M.S. Lew, APR Principles of Visual Information Retrieval, Springer Verlag, 2001.

[14] M.S. Lew, N. Sebe, C. Djeraba and R. Jain - Content-based multimedia information retrieval: State of the art and challenges. ACM Transactions on Multimedia Computing, Communications and Applications, volume 2, issue 1, pages 1-19, february 2006.

[15] J. Malo, J. Gutierrez, I. Epifanio, and F.J. Ferri, “Perceptually weighted optical flow for motion-based segmentation in MPEG4 paradigm”, Electronic Letters, Volume 36 Issue 20, Pages 1693-1694

[16] A.A.J. Oerlemans, M.S. Lew, E.M. Bakker - Detecting and identifying moving objects in real-time. Proc Advanced School for Computing and Imaging conference, The Netherlands, 2005.