Object tracking system using colour histograms*

J. Vergés-Llahí†, J. Aranda§, A. Sanfeliu†.
†Institut de Robòtica i Informàtica Industrial (UPC-CSIC).
§Dept. d'Enginyeria de Sistemes, Automàtica i Informàtica Industrial (UPC).

Abstract

A pan-tilt structure has been endowed with the capability of doing the visual tracking of a moving target. Moving objects in the field of view of the camera are detected and the more relevant region colour feature is selected as the pattern to follow. Colour histogram is used as reliable feature to model object appearance and its adaptation handles with illumination changes.

Keywords: Active vision, visual tracking, real-time image processing.

1 Introduction

In this work we consider tracking moving objects with an active camera. Active vision [1] implies computer vision implemented with a movable camera. One of the most important and basic behaviours of an active camera is to keep a given object (the target) centred in the image.

Two different approaches have been used to perform tracking with an active camera: motion-based and feature-based. Classical motion-based techniques can only be used with a static camera where background remains unchanged. When using a mobile camera, this can be overcome by stopping the tracker to get measures [2] or by means of background compensation [3].

Feature-based methods are based on modified classic pattern recognition algorithms where the target is recognised in successive images and its position is estimated. The major drawback here is that the target has to be recognised for each and every new image. Such operation is usually considered at the top level of image processing both due to data intrinsic complexity and to high computational cost associated to its solution in real time.

---

* Work partially funded by CICYT project TAP98-0473 and Catalonia Government.

Reliability of recognition process and execution time measure tracking systems efficiency. Some solutions improve processing time by using expensive computational resources which limits their generalised application in industry. With standard image acquisition hardware, the main constraint is computation time [4]. Colour is a good feature in order to reduce the amount of data to process without losing the robustness of matching.

In our work, colour histogram has been used instead of any other sort of colour cluster description, as in [5],[6],[7],[8],[9], due to its simplicity, versatility and velocity, needed in tracking applications. Moreover, it has been vastly proven its use in colour object recognition by *colour indexing* [10],[11]. Main colour drawback is its sensitivity to changes that illumination and object motion carry out on object colour. *Constancy colour* techniques try to avoid this bad effects. For example, in [12] is developed a correction for colour object indexing in [10]. Other colour adapting techniques can be found in [5],[7],[9],[13], but we try to adapt our histogram description by means of a Kalman filter.

2 Pan and tilt head vision system

The goal of this system is to extract the necessary information from the image sequence grabbed by the head cam to track an object trajectory. The observer is initially stationary to the scene background and we use a model-based approach to characterize the object to track. The system has two different main parts: *model learning* and *object tracking*. We will also describe object features and how we use the colour histogram technique in this approach.

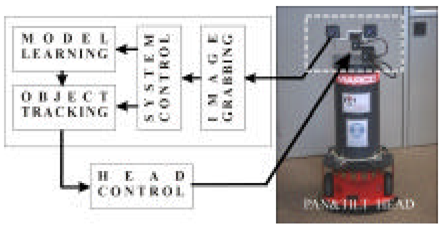

Figure 1: Pan&tilt Head Vision System.

2.1 Object features

Object model has two sorts of features, related to its apparent size in the image –such are area in number of pixels and the size of a window surrounding the object, the window position and that of the object centre of gravity with respect the image– and some more information about how the object looks like, represented by object colour histogram.

Be a colour image $I:[0, width-1]\times[0, height-1]\cap Z^2 \rightarrow [0,255]^3 \cap Z^3$ and $p=(x,y)$ a colour point, then $H=\{H_{ijk}=\#\{p \mid I(p)\in Bin_{ijk}\},(i,j,k)\in[0,N-1]^3\}$ is a colour histogram, where Bin_{ijk} is a colour bin and N is the number of buckets. Then, operations with histograms, such as distance, intersection, union, subtraction, are defined as follows:

$$dist(H,T)=1-\sum_{ijk}\min\{H_{ijk},T_{ijk}\} \qquad (1)$$

$$H\cap T=\{\min\{H_{ijk},T_{ijk}\}\}_{(i,j,k)\in[0,N-1]^3} \qquad (2)$$

$$H\cup T=\{\max\{H_{ijk},T_{ijk}\}\}_{(i,j,k)\in[0,N-1]^3} \qquad (3)$$

$$H\setminus T=\{H_{ijk},\ if\ T_{ijk}=f\}_{(i,j,k)\in[0,N-1]^3} \qquad (4)$$

Model histogram is used to segment images into object component and background component, being able to extract newly object features, updating the model, and computing commands to keep pan&tilt head tracking the object. Instead of applying colour indexing scheme by histogram intersection to find out where the object is, as in [10], the solution we have adopted is very close to colour histogram backprojection technique in [13]: we binarize an image by deciding for every pixel whether it belongs or not to model histogram. In our case, we compute the intersection between a certain neighbourhood $B_e(I(p))$ of the colour of image pixel p and the model histogram H. If $B_e(I(p))\cap H \neq f$, this colour belongs to the model histogram and, therefore, the pixel p belongs to the object.

2.2 Model learning

Model-based tracking algorithms need to learn a model of the target to follow. This learning can be supervised if it is known which object to track, or unsupervised otherwise, tracking then any single mobile object. It is an important issue when to re-learn a model. In supervised learning, since we know *a priori* what to track, a *ground-truth* consisting on a fix model can be established. Then, if the adaptive model gets too far from the ground-truth one, it is resumed to the fixed one. In unsupervised case, the system tracks the biggest mobile object within visual field. Unlike previous case, the system must to detect the target before. When the system loses its target, it goes again to the target detecting step.

In supervised learning, to extract the object colour histogram from a predefined window it is necessary to separate the object from the background. In structured environments, we can do this by eliminating from the window histogram the histogram computed from the image without this window. Area and position features can be extracted by blob analysis over binarized image obtained by colour histogram backprojection.

Unsupervised learning needs a previous step of independent motion detection to focus the mobile object search and to learn the model. The most difficult case is that of mobile observer because of the self-induced background image motion [14]. But, for a initially stationary observer, an easy way to detect independent motion is *frame differencing*, as is in our case. The images of a historical sequence are combined with weights depending on its position within the sequence: the oldest an image is, the least the image weights. This filters small fluctuations and changes in images, and eases to focus on fast changes due to mobile objects[15].

$$I' = \sum_i a_i \cdot I_i \text{, where } a_i = \frac{i}{\sum_i i} \text{ and } \sum_i a_i = 1 \quad (5)$$

The resulting image is then compared with the real image at this step to obtain the difference image. Only if there has been some motion, differentiated regions will appear, which are considered to belongs to the same mobile object. Then, all features can be computed as explained before.

2.3 Object tracking

Object tracking process has three steps: object segmentation (colour histogram back-projection), feature extraction (blob analysis) and, finally, histogram adaptation by means of a Kalman filter to cope with colour variations:

$$\hat{H}_{ijk}^{t+1} = b \cdot \hat{H}_{ijk}^{t} + (1-b) \cdot H_{ijk}^{t},\ b \in [0,1] \quad (6)$$

where H^{t+1} is the predicted histogram value for step *(t+1)*, H^t is the predicted histogram value at step *t* and H^t is the computed histogram value at step *t*. Histogram adaptation takes into account colour changes due to relative motion between observer, object and illuminant.

3 Control module

Tracking module feeds the *head control module* at 10Hz with the current *(x,y)* position of the target in the image. Its goal consists in moving the head to keep the target centred in the image. The error signal, $e = (e_x, e_y)$, to compute head movements is defined as the difference between the target image position *(x,y)* and the image centre pixel coordinates. Head and camera geometry have been considered to calculate pan and tilt motion [16]. Camera base design keeps the camera rotating around its optical centre to simplify head-camera kinematical model, although this is not totally fulfilled in practice due to mechanical constraints and calibration mistakes. Therefore, a proportional controller can be defined as $w = [w_{pan}, w_{tilt}]^T = k_p / f \cdot e(t)$, where k_p is the vector of proportional gains and f is the camera focal length. No integrative or derivative gains have been needed in this implementation.

4 Experimental results

The overall system consists on the pan&tilt unit (PTU-46-17.5 from *Direct Perception* company) with two orthogonal degrees of freedom moved by stepped motors and equipped, a camera and a PC Pentium 400MHz fitted with a standard image acquisition board (Matrox Meteor). The PTU control unit can be driven by position and velocity commands via RS-232. The experimental results presented here, including the image processing and visual tracking, are computed in real time (10 Hz).

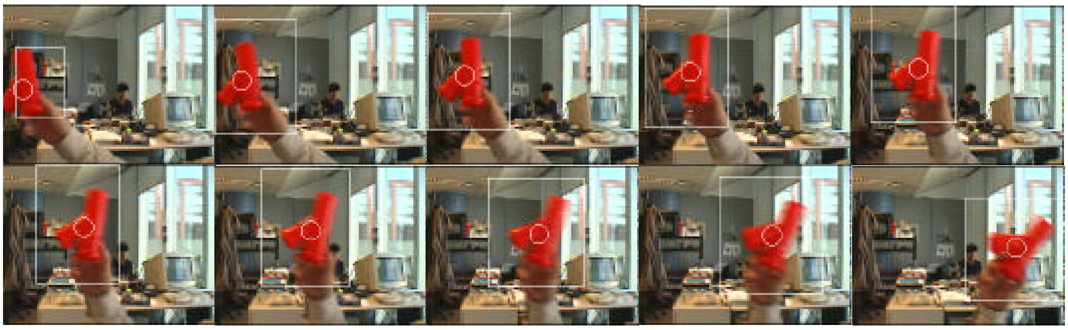

Figure 2. Shows some results with one of these experiments.

5 Conclusions

In this paper it has been shown how can be implemented a low cost real-time pan and tilt tracking system based on colour information for many practical applications, such as robot control feedback, teleoperation, surveillance and security, video-conference systems and home video cameraman. Moreover, it has been shown that colour histograms are also suited to fast object segmentation using colour histogram backprojection and a way to adapt colour information to changes using Kalman filtering.

References

[1] J.Y. Aloimonos, D.P. Tsakiris. On the visual mathematics of tracking. *Image and Vision Computing.* Vol.9(4): 235-251, 1991.

[2] L. Chen, S.Chang. A video tracking system with adaptive predictors. *Pattern Recognition*, Vol. 25(10): 1171-1180, 1992.

[3] D.W. Murray, A. Basu. Motion Tracking with an active camera. *IEEE Trans. on Pattern Analysis and Machine Intelligence,* Vol. 16(5), 1994.

[4] A. Arsenio, J. Santos-Victor. Robust Visual Tracking by an Active Observer. *IROS'97*. 1997.

[5] Daesik Jang, Hyung-Il Choi. Moving object tracking using active models. *ICIP'98*, Vol.3: 648 -652, 1998.

[6] B. Heisele, U. Kressel, W. Ritter. Tracking non-rigid, moving objects based on color cluster flow. *CVPR '97 San Juan*: 253-257, 1997.

[7] S. J. McKenna, Y. Raja, Shaogang Gong: Tracking Colour Objects Using Adaptive Mixture Models, *Image and Vision Computing*, Vol. 17: 225-231, 1999.

[8] T. Nakamura, T. Ogasawara. Online visual learning method for color image segmentation and object tracking. *IROS '99*. Vol. 1: 222 –228, 1999.

[9] R. Schuster. Color object tracking with adaptive modeling. *Workshop on Visual Behaviors, International Association for Pattern Recognition*: 91-96, June 1994.

[10] M.J. Swain, D.H. Ballard. Color Indexing. *IJCV*, Vol. 7(1): 11-32, 1991.

[11] B. Schiele and J. L. Crowley. Object recognition using multidimensional receptive field histograms. *ECCV'96*, Vol I, pages 610-619, 14-16 April 1996.

[12] B.V. Funt, G.D. Finlayson. Color Constant Color Indexing. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 17(5): 522-529, May 1995.

[13] Tse Min Chen, R.C. Luo, Tsu Hung Hsiao. Visual tracking using adaptive color histogram model. *IECON '99*, Vol. 3: 1336 -1341, 1999.

[14] P. Nordlund and T. Uhlin, Closing the Loop: Detection and Pursuit of a Moving Object by a Moving Observer. *Image and Vision Computing*, Vol. 14: 265-275, 1996.

[15] J. Aranda, C. Leon, M. Frigola. A multi-tracking system for trajectory analysis of people in a restricted area. *Time-varying image processing and moving object recognition. Elsevies Science*:118-124, 1994.

[16] J. Aranda, J. Climent, A. Grau. Real –Time pan and tilt tracking of moving targets. *Controlo'98*. 1998.