The Application of Colour Filtering to Real-Time Person Tracking*

N.T. Siebel and S.J. Maybank†

Computational Vision Group
Department of Computer Science
The University of Reading, Whiteknights,
Reading, Berkshire, RG6 6AY, UK

http://www.cvg.cs.reading.ac.uk/
{nts,sjm}@isg.cs.reading.ac.uk

January 2, 2002

Abstract

We present results from multiple experiments with colour filtering methods in order to improve robustness in an integrated surveillance system to track people in a subway station. The system is designed to operate in real-time in a distributed local network of off-the-shelf computers, resulting in practical constraints not found in developmental systems. We show that the quality of colour information is degraded by electrical interference and image compression to such an extent that it is no longer useful for local edge detection. We give a recommendation as to what methods can be used to filter out most of the image noise influencing local edge detection and show how using these methods increases robustness of tracking.

Introduction

With the recent advances of computer technology real-time automated visual surveillance has become a popular area for research and development. Surveillance cameras are installed in many public areas to improve safety, and computer-based image processing is a promising means to handle the image databases generated by large networks of cameras. The task of an integrated surveillance system is to warn an operator when it detects events which may require human intervention, for example to avoid possible accidents. The warnings can only be reliable if the system can detect and understand human behaviour and for this it must locate and track people reliably.

A frequent problem with surveillance systems is the poor quality of the video images. Installed cameras are often old, and images are transmitted down long analogue cables which are vulnerable to interference and signal degradation. There are many unsolved problems when it comes to processing this noisy data by means of a computer.

Many methods to make image analysis more

*A previous version of this article appeared in the *Proceedings of the 2nd European Workshop on Advanced Video-Based Surveillance Systems (AVBS'2001)*, pages 227–234, September 2001.

†This work is funded by the European Union, grants ADVISOR (IST-1999-11287) and Smart II (FMRX-CT96-0052).

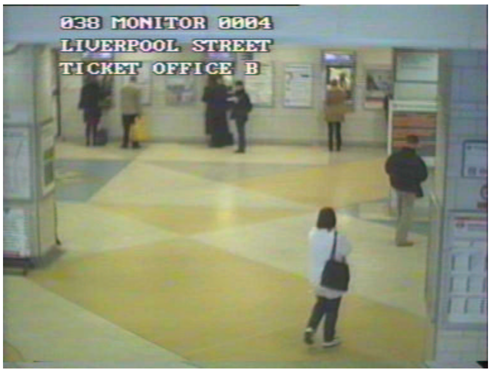

Figure 1: View from surveillance camera

robust involve the introduction of prior knowledge, for example the use of complex statistical models to detect image motion [2], and the use of detailed 3D models of the human body to give more accurate tracking results [4, 7].

However, automated visual surveillance systems have to operate in real-time and with a minimum of hardware requirements, if the system is to be economical and scalable. Even with today's computer speeds this limits the complexity of methods for detection and tracking. Haritaoglu et al [5] show that real-time performance can be obtained with a simplified person model in low-resolution images.

In this paper we use simple image filtering techniques to improve a people tracking method in which the person model is of medium complexity. We explore whether it is possible to use the colour information which is available on many systems today. We show image filtering methods which can enhance edge detection, thereby improving the robustness of image analysis software for people tracking.

1 People Tracking

The tracking system used for our experiments is an extension of the Leeds People Tracker which was developed by Baumberg and Hogg [1]. We have ported the tracker from an `sgi`™ platform to a PC running GNU/Linux to make economic system integration feasible.

The tracker uses an active shape model [1] for the contour of a person in the image. A space of suitable models is learnt in a training stage using a set of video images containing walking pedestrians. Detected person outline shapes are represented using cubic B-splines, giving a large set of points in a high-dimensional parameter space. Principal Component Analysis (PCA) is applied to the obtained set of shapes to generate a lower dimensional subspace S which explains the most significant modes of shape variation, and which is a suitable state space for the tracker.

1.1 Basic detection and tracking

People tracking is performed in multiple stages. The tracker maintains a background image which is automatically updated by median-filtering the sequence of video images over time. To detect new people, a motion detector subtracts the background from the current video image. Thresholding of this difference image yields a binary image containing regions where the pixel values of the current image differ significantly from the pixel values of the background image. The detected foreground regions that match certain criteria for size and shape are examined. Their outline shape is approximated by a B-spline and projected into the PCA space S of trained pedestrian outlines. The new shape obtained in this process is then used as a starting point for further shape analysis. Once people are recognised they are tracked using the trained shape model. This is done with the help of Kalman filtering using second order motion models for their movements. The state of the tracker includes the current outline shape as a point in S, which is updated as the observed outline changes during tracking.

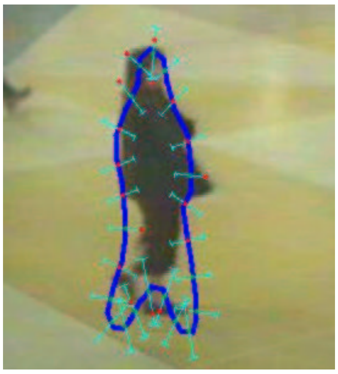

Figure 2: Edge search for shape fitting

1.2 Fitting a person shape

In order to adjust the outline shape to each new image of a person we use an iterative optimisation method. The current estimate of the shape in S is projected onto the image. It is then adjusted to the actual shape in an optimisation loop by searching for edges of the person's outline, at each spline control point around the shape. Figure 2 illustrates this process showing the original camera image, in this case with a fairly dark person against a light background.

It is in this measurement process that image noise and background clutter severely disturb the search for the best fitting outline.

2 Colour Filtering

In order to make edge detection more robust we filter the image in such a way that foreground objects become more salient. Normally the search for edges is done in the pixelwise absolute difference image obtained by subtracting the background image from the video image, to improve contrast [1].

2.1 Image Quality

When dealing with colour information contained in video images, their quality is an important issue. Before interpreting colour information obtained from from image measurements it is important to trace the path the image took, from camera to computer. The colour images used in our system originate from interlaced colour cameras installed in subway stations. They are transferred over long cables and affected by interference from electrical equipment in the station. The image sequence used for the experiments shown was recorded in a London Underground station[1], stored on analogue (S-VHS PAL) video tape which degraded the colour information. The image quality was further reduced by JPEG image compression needed for transmission over a local network (Ethernet) in the integrated surveillance system. Even though the human eye does not notice a lot of image degradation, the JPEG compression can create considerable problems for image processing algorithms. The reason is that JPEG compression techniques usually sub-sample colour information (the 2 "chrominance channels", denoted C_B and C_R) [6], thereby causing problems in algorithms which rely on the correctness of this colour information. JPEG stores the image brightness ("luminance", Y) at a higher resolution than chrominance because this component makes up the main part of the information extracted from images by humans.

2.2 Filtering Techniques

Our investigations involved a large number of colour filtering techniques to find out how they influence edge contrast and how they behave in the presence of the image noise described earlier. The filters used project each pixel in the image from the 3-dimensional RGB colour space down to a 1-dimensional subspace, giving a grey level image (see below for exam-

[1]Thanks to London Underground Ltd. and to Sergio Velastin (King's College London) for providing the video recordings.

ples). The resulting image can then be used by our edge detection algorithm to search for those edges included in the outlines of people. We have applied both linear (weighted sums of RGB values) and non-linear mappings (for example, mapping to the hue component in HSV space). However, given the low colour resolution in our images some filters are not feasible in our system. In the following examinations we focus on the results obtained from mappings that are least prone to colour noise.

Another problem for edge detection algorithms is the low contrast of people against the background. The image in Figure 1, chosen as an example, shows a person with a white jacket against a light yellow background. In the colour difference image the jacket appears as dark blue and most edge detection methods will not be able to detect it as the edge contrast at the jacket outline is too low.

2.3 Filtering the difference image

Many algorithms for colour edge detection start by taking the absolute difference in each colour channel to get a new colour image. Each pixel value $p = (R, G, B)$ of this difference is then mapped to a scalar using a function of its components. The mappings use a scaled Euclidian norm, $||p||_2 = \sqrt{\frac{1}{3}(R^2 + G^2 + B^2)}$, or a weighted sum, $w_1 R + w_2 G + w_3 B$ with $\sum w_i = 1$.

The image in Figure 3 shows the result for a simple linear mapping from RGB space onto the luminance (Y) subspace, using the Y projection from CIE[2] recommendation 709, denoted here as Y_{709} (taken from [3]):

$$Y := Y_{709} = .2125\,R + .7154\,G + .0721\,B. \quad (1)$$

The weights are chosen in order to mimic the way humans perceive the brightness of a colour image. However, this weighting only

Figure 3: Luminance Y of colour difference

gave a small improvement for edge detection compared to the simpler method of weighting each colour channel equally.

Looking at the resulting image, one can see that detection of the outline of the lady in the front fails because her coat has too low a contrast with respect to the background. For the person to the right, the main parts of his body, especially the trousers, are many different shades of grey. Searching for edges around his shape outline is very sensitive to the parameters used in edge detection, which can cause difficulties and mis-detection.

2.4 Differencing the filtered images

In a second stage of experiments we filtered both the background image and the current video image using the same Y projection (1) and subtracted the results. You can see the difference image in Figure 4.

Compared to the previous image there is not much change when it comes to the woman's coat. Other parts of both people in the front, however, appear much more salient, standing out clearly against the background. For edge detection this gives a significant advantage because it increases robustness by being less dependent on detection parameters such as threshold values. Looking at the right hand

[2] Commission Internationale de L'Éclairage

Figure 4: Absolute difference of luminance Y

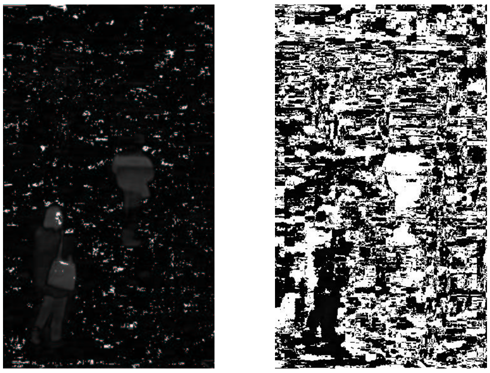

Figure 5: Chrominance channels C_B and C_R

person's coat and trousers we see that these are now more uniformly shaded which means that edge detection is also less prone to being distracted by edges located inside the body.

This method of filtering also makes background edges stand out more clearly. For motion detection algorithms this could create difficulties. Our edge detector, however, operates locally in the image so this does not pose a problem.

3 Discussion

In this paper we have presented an approach to improve local edge detection for people tracking systems in order to increase robustness in the presence of image noise. The main problem in an integrated system like ours is the noise found in the chrominance channels of the image as seen in Figure 5. This is due to electrical interference in a subway station (analogue camera cabling) and image compression techniques (such as JPEG) within our local computer network.

We have shown that using image differencing methods using only luminance gives better results in local edge detection than standard routines, thereby improving local edge detection which is a crucial part of our people detection and tracking system.

It should be noted that even in a system like ours, colour information can still be of good use for the identification of people once they are tracked. Examining the influence of strong colour noise on the identification of tracked people is part of our ongoing research.

References

[1] Adam M. Baumberg. *Learning Deformable Models for Tracking Human Motion.* PhD thesis, School of Computer Studies, University of Leeds, October 1995.

[2] Ahmed Elgammal, David Harwood, and Larry Davis. Non-parametric model for background subtraction. In David Vernon, editor, *ECCV 2000, 6th European Conference on Computer Vision*, pages 751–767. Springer Verlag, 2000.

[3] Adrian Ford and Alan Roberts. *Colour Space Conversions*, August 1998.

[4] Dariu M. Gavrila and Larry S. Davis. Tracking of humans in action: A 3-D model-based approach. In *ARPA Image Understanding Workshop*, pages 737–746, Palm Springs, February 1996.

[5] Ismail Haritaoglu, David Harwood, and Larry S. Davis. W^4: Real-time surveillance

of people and their actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):809–830, August 2000.

[6] Thomas G. Lane. *IJG JPEG Library: System Architecture*. Independent JPEG Group, 1991–1995. Part of the Independent JPEG Group's JPEG software documentation.

[7] Hedvig Sidenbladh, Michael J. Black, and David J. Fleet. Stochastic tracking of 3D human figures using 2D image motion. In David Vernon, editor, *ECCV 2000, 6th European Conference on Computer Vision*, pages 702–718. Springer Verlag, 2000.