ATTENTION-BASED VANISHING POINT DETECTION

Fred Stentiford
University College London, Adastral Park Campus, Martlesham Heath, Ipswich, UK
f.stentiford@adastral.ucl.ac.uk

ABSTRACT

Perspective is a fundamental structure that is found to some extent in most images that reflect 3D structure. It is thought to be an important factor in the human visual system for obtaining understanding and extracting semantics from visual material. This paper describes a method of detecting vanishing points in images that does not require prior assumptions about the image being analysed. It enables 3D information to be inferred from 2D images. The approach is derived from earlier work on visual attention that identifies salient regions and translational symmetries.

Index Terms— Machine vision

1. INTRODUCTION

Perspective is present to some extent in all images that reflect 3D information. Parallel lines in the three dimensional scene project to vanishing points in the image. Locating the vanishing points provides a powerful way of inferring 3D structure from a 2D image especially in a man-made environment where a scene may be captured into an architectural CAD program, for example. Lutton et al [1] apply the Hough transform to assemble line segments which point towards vanishing points. This approach requires knowledge of camera parameters and relies heavily upon edge extraction to identify relevant features. Problems arise when dealing with large numbers of very short segments that arise in certain images. McLean et al [2] cluster gradient orientations to again detect line structure in images and evaluates the method against two grey level images. Along with other authors Shufelt [3] uses a Gaussian sphere representation and addresses the problem of spurious edges in images with a limited range of object orientations.

Rother [4] applies the ideas to architectural environments and rejects falsely detected vanishing points by making use of camera parameters. Cantoni et al [5] explores two approaches, one using the Hough transform and the other edge detection. Successive analyses are required to locate multiple vanishing points. Almansa et al [6] proposes a method not dependent on camera parameters which searches for image regions that contain maximum numbers of line segment intersections. Curved boundaries in images that do not contain actual vanishing points can cause false alarms. Gabor wavelet filters are used by Rasmussen [7] to obtain dominant texture orientation in images of roads.

The approach taken in this paper is based upon a model of human visual attention [8,9] that identifies what is important in a scene. The same basic mechanism has been used to extract reflective symmetries [10] in images, correct the colour balance in poorly illuminated photos [11], and measure similarity [12]. The approach bears some similarity with the RANSAC algorithm in that the goal is sought by iteratively selecting random subsets of data points. However, whereas RANSAC normally requires a pre-defined and parameterised model to which points are fitted, no such model is assumed here. The next sections briefly outline this model and how it is modified to extract symmetries of perspective. Some illustrative results on natural images are provided.

2. VISUAL ATTENTION

Salient regions in images may be detected through a process that compares small regions with others within the image. A region that does not match most other regions in the image is very likely to be anomalous and will stand out as foreground material. For example, the edges of large objects and the whole of small objects normally attract high attention scores mainly because of colour adjacencies or textures that only occur rarely in the image. Repetitive backgrounds that display a translational symmetry are also assigned low attention scores.

Region matching requires a few pixels (a fork) within that region to match in a translated position in another region. If the difference in colour of one pixel pair exceeds a certain threshold a mismatch is counted and the attention score is incremented.

Let a pixel $\boldsymbol{x}$ in an image correspond to a measurement $\boldsymbol{a}$ where

$$\boldsymbol{x} = (x_1, x_2) \text{ and } \boldsymbol{a} = (a_1, a_2, a_3)$$

Define a function $\boldsymbol{F}$ such that $\boldsymbol{a} = \boldsymbol{F}(\boldsymbol{x})$.

Consider a neighbourhood N of $\boldsymbol{x}$ with radius $\boldsymbol{r}$ where

$$\{\boldsymbol{x}' \in N \text{ iff } |x_i - x'_i| < r_i \ \forall i\}$$

Select a fork of m random points S_x in N where

$$S_x = \{\boldsymbol{x}'_1, \boldsymbol{x}'_2, \boldsymbol{x}'_3, ..., \boldsymbol{x}'_m\}$$

Shift S_x by a displacement $\boldsymbol{\delta}$ in the image to become S_y where

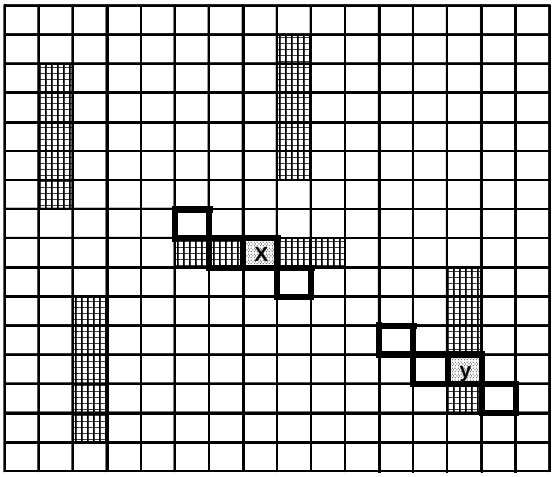

Figure 1. Fork at x mismatching at y with $\boldsymbol{\delta} = (6,4)$.

$S_y = \{\boldsymbol{x}'_1 + \boldsymbol{\delta}, \boldsymbol{x}'_2 + \boldsymbol{\delta}, ..., \boldsymbol{x}'_m + \boldsymbol{\delta}\}$ and $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{\delta}$ The fork S_x matches S_y if

$$|F_j(\boldsymbol{x}'_i) - F_j(\boldsymbol{x}'_i + \boldsymbol{\delta}_i)| < \varepsilon_j \ \forall\ i,j\ .$$

In Figure 1 a fork of $m = 4$ pixels $\boldsymbol{x}'$ is selected in the neighbourhood of a pixel $\boldsymbol{x}$ and is shown mismatching in the neighbourhood of pixel $\boldsymbol{y}$. The neighbourhood of the second pixel $\boldsymbol{y}$ matches the first if the colour intensities of the corresponding pixels all have values within $\boldsymbol{\varepsilon}$ of each other. The attention score $V(\boldsymbol{x})$ for each pixel $\boldsymbol{x}$ is incremented each time a mismatch occurs in the fork comparisons with a sequence of pixels $\boldsymbol{y}$. A location $\boldsymbol{x}$ will be worthy of attention if a sequence of t forks matches only a few other neighbourhoods in the space. Pixels $\boldsymbol{x}$ that achieve high mismatching scores over a range of t forks S_x and pixels $\boldsymbol{y}$ are thereby assigned a high estimate of visual attention. An application to image compression is described in [9]. It should be noted that this technique takes no account of saliency which might arise from semantic relationships with other images.

3. VANISHING POINT DETECTION

In this paper measures of perspective are computed using the same mechanism for measuring attention, except that forks are passed through a perspective transform *before* translation and testing for a match. Peaks in the distributions of matches across the image indicate the locations of vanishing points in the image. Forks are constrained to include some (h) pixels that mismatch each other. This ensures that the measure is directed towards attentive regions of the image and not large background tracts of self-matching sky, for example.

A fork of m random pixels S_x is defined as a set of pixel positions where

$$S_x = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, ..., \boldsymbol{x}_m\}\ .$$

A series of M such forks is given by

$$S_x^k = \{\boldsymbol{x}_{1k}, \boldsymbol{x}_{2k}, \boldsymbol{x}_{3k}, ..., \boldsymbol{x}_{mk}\} \quad k = 1,2,\ldots, M \quad (1)$$

with

$$|F_j(\boldsymbol{x}_{pk}) - F_j(\boldsymbol{x}_{qk})| > \varepsilon_j \text{ for at least } h \text{ values of } p.$$

Transformed forks S_y^k are generated by transforming the S_x^k where

$$S_y^k = \{\boldsymbol{y}_{1k}, \boldsymbol{y}_{2k}, \boldsymbol{y}_{3k}, ..., \boldsymbol{y}_{mk}\}$$

with $$\boldsymbol{x}_0 - \boldsymbol{y}_{ik} = \alpha\,[\,\boldsymbol{x}_0 - \boldsymbol{x}_{ik}]\ \forall\ i,k \quad (2)$$

where α is a scalar constant and $\boldsymbol{x}_0$ is the location of a potential vanishing point.

The fork S_y^k is now a perspective transformed and shifted version of S_x^k, and matches the pixels in S_x^k indicating a possible vanishing point at $\boldsymbol{x}_0$ if

$$|F_j(\boldsymbol{x}_{ik}) - F_j(\boldsymbol{y}_{ik})| < \varepsilon_j\ \forall\ i,j.$$

Figure 2 shows a 5 pixel fork matching an image and its perspective transform also matching the image thereby providing evidence for a possible vanishing point at the dot. In this case transformed versions of all forks containing just white pixels would trivially match the background and are excluded by (1). Distributions of measures of the strength of perspective across the image are produced in the following steps:

1. Set histogram of perspective values to zero.
2. For each image pixel (i,j) ($\boldsymbol{x}_0$ in (2))
3. Generate a fork S_x^k with h pixels mismatching remaining $(m-h)$ pixels and including pixel (i,j)
4. Transform S_x^k with $\alpha = 1/2$.
5. If S_y^k matches increment histogram bin at (i,j).
6. Loop to step 3 k = M times.

Loop to step 2 for each pixel.

4. RESULTS

A number of images from the Corel Database with obvious perspective structure were processed and the measure of perspective at each pixel calculated as above. In these results, except where otherwise stated, the number of elements (m) in each fork was set at 12, the number of comparisons (M) at 100, and α at 0.5. Performance was not very sensitive to α, however values close to unity or zero gave noisy results and a middle value was used. The position of the peak marking the principal vanishing point was indicated on the image and the individual scores plotted as 3D histograms.

In Figure 3 the perspective of the road, trees and sky all appear to converge on virtually the same point. The distribution of highest scores centre on a maximum value in this same area. Figure 4 shows the distribution of scores generated using single and 3 pixel forks. It is apparent that the scores become less noisy and are more peaked at the vanishing point as the number of fork pixels

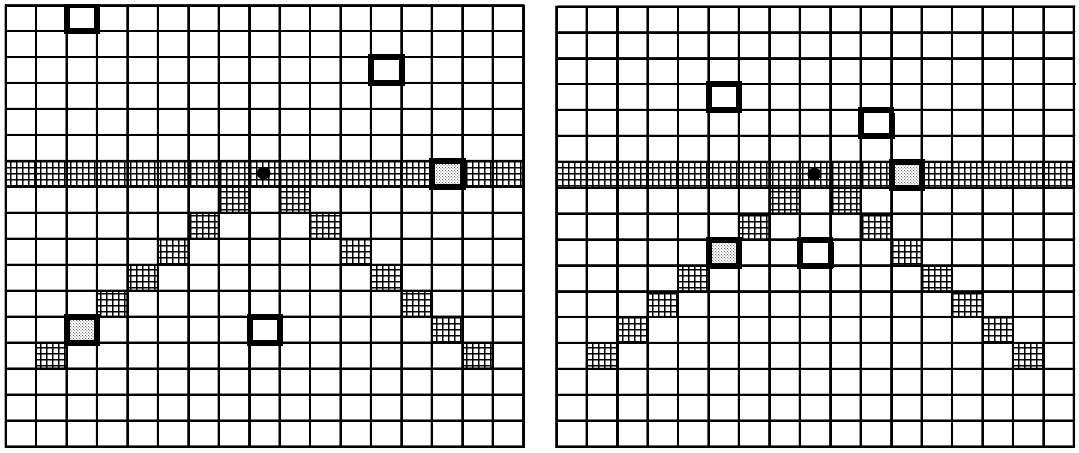

Figure 2. Pixels transformed with $\alpha = 0.5$ and matching image.

is increased. The peak of perspective is slightly to the left of the entrance in Figure 5 because of the asymmetric vegetation. Subsidiary peaks follow the lines of the hedges on each side. The detected vanishing point in Figure 6 lies on the horizon but slight asymmetry pulls the vanishing point to the right of that which might be indicated by the road. In Figure 7 the same vanishing point is obtained on the extreme right using just a part of the image in Figure 6. Figure 8 highlights those pixels that play a part in the matching during the calculation of scores in the vicinity of the vanishing point in Figure 6. It can be seen that most of the image apart from certain portions of the sky and the path contribute to the score. High scoring pixels cover the neighbourhood of the road as it disappears to the left in Figure 9. The peak takes into account the trees and the sky as well as the road and its markings. Finally the principal vanishing point in figure 10 is identified.

5. DISCUSSION

More accurate results with larger forks could indicate that features containing most perspective information span the whole image rather than more local neighbourhoods.

The results generated in this paper restrict the location of vanishing points to within the boundaries of the image, but the approach is equally applicable to the detection of vanishing points outside the image providing the matching fork pixels S_x^k and S_y^k themselves all lie within the image. In this special case of testing for vanishing points at infinity the transform becomes $y_{ik} = x_{ik} + \delta_{ik}$ and peaks in the distribution of the direction of the shifts δ_{ik} for matching forks give the directions of the distant vanishing points.

Key advantages in this approach over other techniques include the absence of the need for the specification of any a priori features that might characterise the presence of perspective, such as edges or resonance with specific types of filter. This method should still function even if the image is blurred and contains no sharp edges. In addition no knowledge is required of camera parameters or their calibration and no restrictions are placed on the minimum strength of any perspective structure that must be present in the data for the algorithm to function effectively. Finally there is no manual intervention necessary to either initialise or guide the process. However, further work is clearly necessary to assess the performance on much larger sets of data.

The results reported in this paper have been produced with 100 iterations of fork generation per pixel. Although the computational steps are very simple there are a large number of them and a vanishing point analysis takes about 10 seconds on a 1.8GHz machine running in C++. The computation may be reduced on a sequential machine by only scoring sampled pixels where it may not be necessary to obtain positional accuracy to the nearest pixel. However, the matching of forks can be carried out in parallel as each match is independent of the next and related implementations on the Texas Instruments DM642 DSP platform indicate that processing can take place at video speeds.

6. CONCLUSIONS

This paper has described a technique for extracting vanishing points in images that does not require manual intervention or the prior specification of features that characterise perspective. The features or forks are produced through a modified attention focussing mechanism that selects the best positions that maximise the matching of forks after a perspective transform. Future work will be to test the approach on a greater diversity of images and to implement the algorithm at video speeds with a view to capturing scenes for virtual worlds.

This research has been conducted with the support of British Telecom and within the framework of the

European Commission funded Network of Excellence "Multimedia Understanding through Semantics, Computation and Learning" (MUSCLE) [13].

11. REFERENCES

[1] E. Lutton, H. Maitre, and J. Lopez-Krahe, "Contribution to the determination of vanishing points using Hough transform," IEEE Trans. on PAMI, vol. 16, no. 4, pp 430-438, 1994.
[2] G.F. McLean, and D. Kotturi, "Vanishing point detection by line clustering," IEEE Trans. on PAMI, vol. 17, no. 11, pp 1090-1095, 1995.
[3] J.A. Shufelt, "Performance evaluation and analysis of vanishing point detection techniques," IEEE Trans. on PAMI, vol. 21, no. 3, pp 282-288, 1999.
[4] C. Rother, "A new approach for vanishing point detection in architectural environments," 11th British Machine Vision Conference, Bristol, UK, September, 2000.
[5] Cantoni, V., Lombardi, L., Porta, M., and Sicard, N., "Vanishing point detection: representation analysis and new approaches," 11th Int. Conf. on Image Analysis and Processing, Palermo, Italy, September, 2001.
[6] A. Almansa, and A. Desolneux, "Vanishing point detection without any a priori information" IEEE Trans. on PAMI, vol. 25, no. 4, pp 502-506, 2003.
[7] C. Rasmussen, "Texture-based vanishing point voting for road shape estimation," British Machine Vision Conference, Kingston, UK, September, 2004.
[8] F.W.M. Stentiford, "Automatic identification of regions of interest with application to the quantification of DNA damage in cells," Proc. SPIE. Vol. 4662, pp 244-253, 2002.
[9] F.W.M Stentiford, "An estimator for visual attention through competitive novelty with application to image compression," Picture Coding Symposium, Seoul, pp101-104, 2001.
[10] F.W.M. Stentiford, "Attention based symmetry detection in Colour Images," MMSP, Shanghai, October 30-November 2, 2005.
[11] F.W.M. Stentiford, "Attention based colour correction," SPIE Human Vision and Electronic Imaging XI, San Jose, Jan. 2006.
[12] F.W.M. Stentiford, "Attention based similarity," Pattern Recognition, in press, 2006.
[13] Multimedia Understanding through Semantics, Computation and Learning, Network of Excellence. EC 6th Framework Programme. FP6-507752. http://www.muscle-noe.org/

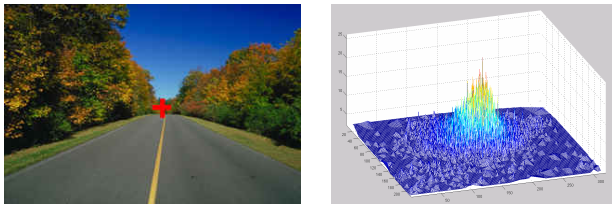
Figure 3.

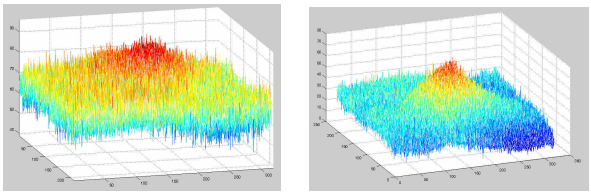
Figure 4.

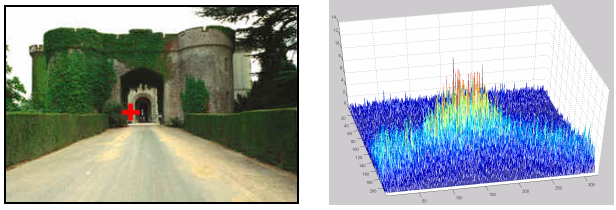
Figure 5.

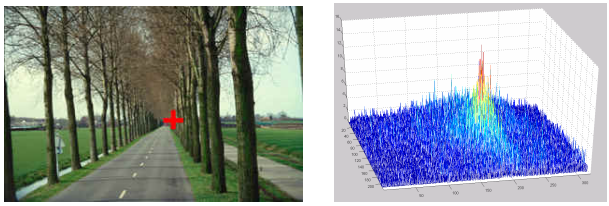
Figure 6.

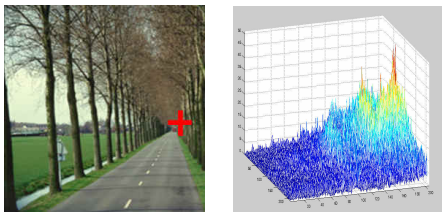
Figure 7.


Figure 8.

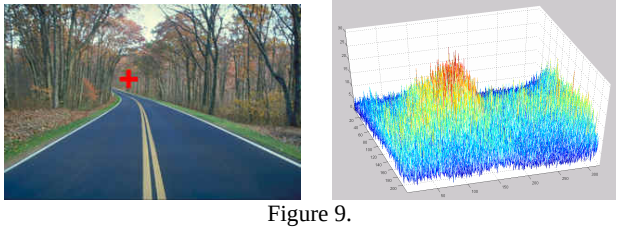
Figure 9.

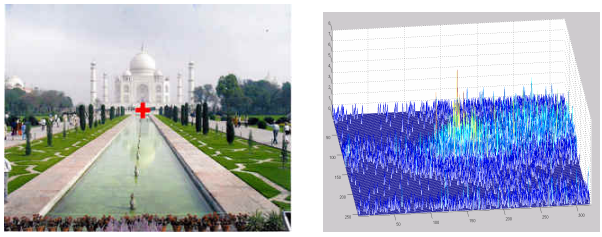
Figure 10.