MATCHING WITH EPIPOLAR GRADIENT FEATURES AND EDGE TRANSFER

Étienne Vincent and Robert Laganière

School of Information Technology and Engineering
University of Ottawa, Ottawa, Canada, K1N 6N5
evincent,laganier@site.uottawa.ca

ABSTRACT

A method for quickly and reliably selecting and matching points from three views of a scene is presented. The points that are selected are based on the concept of epipolar gradients, and consist in stable and relevant image features. Then, the selected points are matched using edge transfer, resulting in a measure of consistency for point triplets and the edges on which they lie, with the camera system's trinocular geometry. This matching scheme is invariant to image deformations due to changes in viewpoint.

1. INTRODUCTION

In computer vision, several applications require to match feature points from a few fixed cameras. To achieve this goal, cameras are first calibrated, then, guided matching is performed. In applications such as telerobotics, where the environment being modelled is continuously changing, these operations must also be fast to allow a continuous update of the match set, from which 3D information is extracted. The reliability of the matches is also crucial, as the integrity of the task can be compromised by mismatches.

The most common approach to sparse matching, when camera calibration is not available, is based on matching Harris feature points, using some measure based on correlation [4, 7, 10]. In [9], a system is presented for fast calibrated matching based on this approach. However, although Harris feature points are relatively stable and fast to compute, it was found that on arbitrary scenes that are not chosen to contain suitable textural content, the detected feature points might not be well distributed or stable enough to completely cover the scene. In the context of 3D reconstruction, this means that important 3D information about scene objects might be missing.

An alternative feature detector is presented in this paper, which is fast and results in more an better distributed matches when camera systems are calibrated. This detector relies on epipolar gradients, an idea introduced in this paper. A simple way to match points is also presented which is appropriate for calibrated triplets of images and is invariant to deformations due to changes in viewpoint.

The next Section quickly reviews some features of trinocular geometry. Then, Section 3 discusses how feature points are chosen. Next, Section 4 describes how feature points are matched, Section 5 describes a constraint which weeds out the possible remaining false matches, and finally, Section 6 presents experimental results.

2. TRINOCULAR GEOMETRY

Two views of a scene are related by the well known epipolar geometry, which can be represented by a 3×3 fundamental matrix F. For a point $\mathbf{x}$ in the first image, the corresponding point in the second image will lie on the epipolar line $\mathbf{l}'$. This line can be obtained from $F\mathbf{x} = \mathbf{l}'$, where $\mathbf{x}$ and $\mathbf{l}'$ are represented in homogeneous coordinates.

When three viewpoints are used, more relationships can be obtained from trinocular geometry. The $3 \times 3 \times 3$ trifocal tensor T relates points $\mathbf{x}$, $\mathbf{x}'$ and $\mathbf{x}''$ in three images:

$$\mathbf{x}''_l = \mathbf{x}'_i \sum_{k=1}^{3} \mathbf{x}_k T_{kjl} - \mathbf{x}'_j \sum_{k=1}^{3} \mathbf{x}_k T_{kil} \quad (1)$$

This define 9 trilinearities for $i, j \in \{1, 2, 3\}$, 4 of which are linearly independent. Although there are more direct ways of estimating the position of a point $\mathbf{x}''$ from its corresponding points $\mathbf{x}$ and $\mathbf{x}'$, more stable results are obtained by solving this over-constrained system of equations.

Trifocal geometry is a powerful tool for matching. Indeed, for a point $\mathbf{x}$ in a first image, the search for correspondence is restricted to the line $\mathbf{l}'$ in the second image. And once the correspondence $(\mathbf{x}, \mathbf{x}')$ is known, the matching point in a third image is determined by the trifocal tensor. In practice, a third image can therefore be used to validate matches between two other images. Once a match is found, it is verified by *transferring* it to the third image using equation (1), and evaluating a similarity measure between the transferred point and the other two [4, 7, 9].

Trifocal tensors can also relate lines between images. Thus, when the equation of a line is known in two images, it can be *transferred* to another one using:

$$\mathbf{l}_i = \mathbf{l}'_j \mathbf{l}''_k T_{ijk} \tag{2}$$

This will be used in Section 4 towards matching feature points using the direction of the edges on which they lie.

3. EPIPOLAR GRADIENT FEATURES

In many applications, it is too costly to compare points in a first image with all points along their epipolar lines to search for matches. Thus, matching is often limited to selected feature points. Feature points in the first image are only compared to feature points along their epipolar line in the second image. There is no need for selecting feature points in the third image, however, as the trinocular geometry limits the search to a single point neighborhood.

Good feature points are those that are likely to be easily distinguishable from each other, and which can be identified robustly, with respect to changes in viewpoint. These points should also, as much as possible, represent significant scene features, to result in a good model of the environment.

The most commonly used such feature points are Harris corners [2]. These are high curvature points on image edges, and were shown to be relatively stable [6]. However, when some scene regions do not contain textures with clear corners, few matching points might be detected. Thus, corners might not be distributed well enough to allow a satisfying reconstruction of the scene using only matched points. Epipolar gradient features overcome these problems.

Let $\mathbf{I}$ and $\mathbf{I}'$ be two images, with $\mathbf{x}$ a point in $\mathbf{I}$, and $\mathbf{x}'$ its corresponding point in $\mathbf{I}'$. Then, $\mathbf{x}'$ will lie on $\mathbf{l}'$, the epipolar line of $\mathbf{x}$ in $\mathbf{I}'$, and similarly, $\mathbf{x}$ will lie on $\mathbf{l}$. Now $\mathbf{l}$ and $\mathbf{l}'$ should also correspond, so all points on $\mathbf{l}$ will have their corresponding point lying on $\mathbf{l}'$. Thus if $\mathbf{X}$, the world point projected onto $\mathbf{x}$ and $\mathbf{x}'$, is centered on a locally planar surface, the points on $\mathbf{l}$ that are immediately next to $\mathbf{x}$ should correspond to points in $\mathbf{I}'$ that lie on $\mathbf{l}'$ and are immediately next to $\mathbf{x}'$. Consequently, the intensity gradient of $\mathbf{I}$, at $\mathbf{x}$, in the direction of $\mathbf{l}$, should be similar to the intensity gradient of $\mathbf{I}'$, at $\mathbf{x}'$, in the direction of $\mathbf{l}'$.

The intensity gradient in the direction of the epipolar line will be referred to as the *epipolar gradient*. It can be computed by projecting $\bigtriangledown\mathbf{I}(\mathbf{x})$ onto $\mathbf{l} = (l_1, l_2, l_3)$, giving the explicit formula:

$$\bigtriangledown_{ep}(\mathbf{x}) = \frac{\bigtriangledown\mathbf{I}(\mathbf{x}) \cdot (\frac{-l_3}{l_1}, \frac{l_3}{l_2})}{\|(\frac{-l_3}{l_1}, \frac{l_3}{l_2})\|} \tag{3}$$

where $\mathbf{l}$ can be obtained using an arbitrary line $\mathbf{k}'$ not going through the second image's epipole as:

$$\mathbf{l} = F^T[\mathbf{k}']_\times F\mathbf{x} \tag{4}$$

Thus, in a pair of images for which the epipolar geometry is known, a point having a high absolute epipolar gradient in one image should have a high absolute epipolar gradient in the other as well. This is why epipolar gradient features are good candidates for matching. Additionally, these points are usually found on the border of significant scene features and are thus more relevant for scene reconstruction.

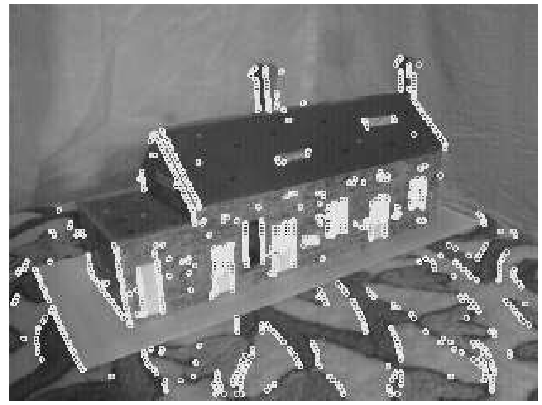

Fig. 1. Detected epipolar gradient feature points

Fig. 1 shows some detected points[1]. Note that points may be detected only on every few lines to limit their number.

4. MATCHING BASED ON EDGE TRANSFER

Now that feature points suitable for matching have been selected, these points must be matched. A common way of comparing potentially matching points is normalized correlation. Such a correlation based approach can give good results when the difference between viewpoints is limited, but is not an accurate measure of similarity in the case of more widely separated views. Then, a measure which is invariant to the reprojection deformation of the area around feature points is needed.

Many such invariant measures have been proposed, notably [1, 3, 5, 8], but they are only invariant to rotation or affine transformations of point neighborhoods, and some are computationally rather expensive. Since here, matching is guided by the camera system's trifocal geometry, points have a few candidate matches, so a more invariant, but less discriminant comparison measure can be used.

Two simple descriptors are used, together with a similarity measure defined between them. The most important descriptor is based on the transfer of lines tangent to edges going through the points in the first two images to the third one using equation (2). These transferred lines should also be tangents to edges going through the corresponding points. Thus, a measure of similarity between three points is the difference between the orientation of the tangent to the edge

[1]obtained from the model house image sequence available at http://www.robots.ox.ac.uk/ vgg/data/

of one of the points, and the orientation of the line obtained by transferring the tangents to edges of the other two.

The other descriptor is simply the intensity value in the image at the pixel. This value should be preserved in different views of the same point taken simultaneously. By itself it is not very discriminating, but it does improve the results from using only edge transfer. Of course, the intensity values near edges are unstable, so the average of neighboring intensities weighted by a gaussian is used.

Let $\Delta I(\mathbf{x}, \mathbf{x}', \mathbf{x}'')$ be the maximum difference between the intensities of $\mathbf{x}$,$\mathbf{x}'$ or $\mathbf{x}''$, and $\Delta\theta(\mathbf{x}, \mathbf{x}', \mathbf{x}'')$ be the difference between the gradient orientation at $\mathbf{x}''$ measured in $\mathbf{I}''$ and computed from the gradients at $\mathbf{x}$ and $\mathbf{x}'$. Then the similarity measure between $\mathbf{x}$, $\mathbf{x}'$ and $\mathbf{x}''$ is:

$$s(\mathbf{x}, \mathbf{x}', \mathbf{x}'') = \max\left(\frac{\Delta I(\mathbf{x}, \mathbf{x}', \mathbf{x}'')}{\sigma_{\Delta\theta}}, \frac{\Delta\theta(\mathbf{x}, \mathbf{x}', \mathbf{x}'')}{\sigma_{\Delta I}}\right) \quad (5)$$

where $\sigma_{\Delta I}$ and $\sigma_{\Delta\theta}$, the respective standard deviations were used to normalize the descriptors to a similar range. This measure will have a low value for corresponding points.

5. DISPARITY CONSISTENCY CONSTRAINT

Sometimes, the similarity measure presented in the previous section might not be discriminating enough. Consequently, even when the search for matches is guided by the trinocular geometry, mismatches can be expected. However mismatches are very undesirable when the goal is reconstruction. Fortunately, when many matches are identified throughout the images, and mismatches are relatively few, they can be eliminated by simply enforcing that matches which are near each other have similar disparities.

To this end, a constraint on the disparity gradient is applied as in [9]. The disparity gradient is a measure of the compatibility of two matches. It is essentially the norm of the difference of the disparities, normalized by the distance between the matches. For two pairs $(\mathbf{x}, \mathbf{x}')$ and $(\mathbf{y}, \mathbf{y}')$, having disparities $d(\mathbf{x}, \mathbf{x}')$ and $d(\mathbf{y}, \mathbf{y}')$ respectively, and their disparity gradient is defined as:

$$\Delta d(\mathbf{x}, \mathbf{x}'; \mathbf{y}, \mathbf{y}') = \frac{|d(\mathbf{x}, \mathbf{x}') - d(\mathbf{y}, \mathbf{y}')|}{|d_{cs}(\mathbf{x}, \mathbf{x}'; \mathbf{y}, \mathbf{y}')|} \quad (6)$$

Where $d_{cs}(\mathbf{x}, \mathbf{x}'; \mathbf{y}, \mathbf{y}')$ is the distance between the midpoint of the disparity vectors. A pairs is considered a mismatch when its disparity gradients with many of its closest neighbors are too high. This eliminates false matches as long as they are not surrounded only by similar false matches, an unlikely situation.

6. EXPERIMENTAL RESULTS

Fig. 3 shows the result of applying the matching scheme to a triplet of images. In the first image, the lines join the coordinate of feature points there, to their coordinate in the second image, and thus represent the disparity between the first two views. Similarly, the lines in the second image indicate the disparity between that image and the third one. Fig. 2 shows the disparities between the first and second images when a Harris detector and correlation are used instead. The same number of feature points were used in both experiments, and the thresholds relevant to the matching process were chosen empirically to maximize the resulting number of matches. It can be seen that the first method obtained more matches (479 versus 414), and provides scene features which are more relevant to scene reconstruction (the matches obtained through the Harris detector being mostly located on the front wall of the house).

Fig. 2. Matched points with Harris features

Fig. 4 shows matches found using the epipolar gradient and edge transfer for simple images of a few objects. Disparities between the first and second images are also shown. Here, 318 matches were found. With the same number of feature points, the Harris detector with correlation only found 31. The proposed method gave significantly better results since these images contain few clear corners, and the difference between their viewpoints is significant.

7. CONCLUSION

In summary, two new techniques were introduced for fast and reliable calibrated sparse matching. A new feature was used based on epipolar gradients, and a new correspondence measure was introduced which relies on transferring edges.

These new techniques are improvements over other approaches. The features based on epipolar gradients are more stable, constitute features which are more relevant to the structure of scenes, and are usually well distributed over images. Matching based on edge direction is fast, and viewpoint independent. Beyond calibrated sparse matching, we believe that epipolar gradients and edge transfer are interesting concepts susceptible of finding other applications.

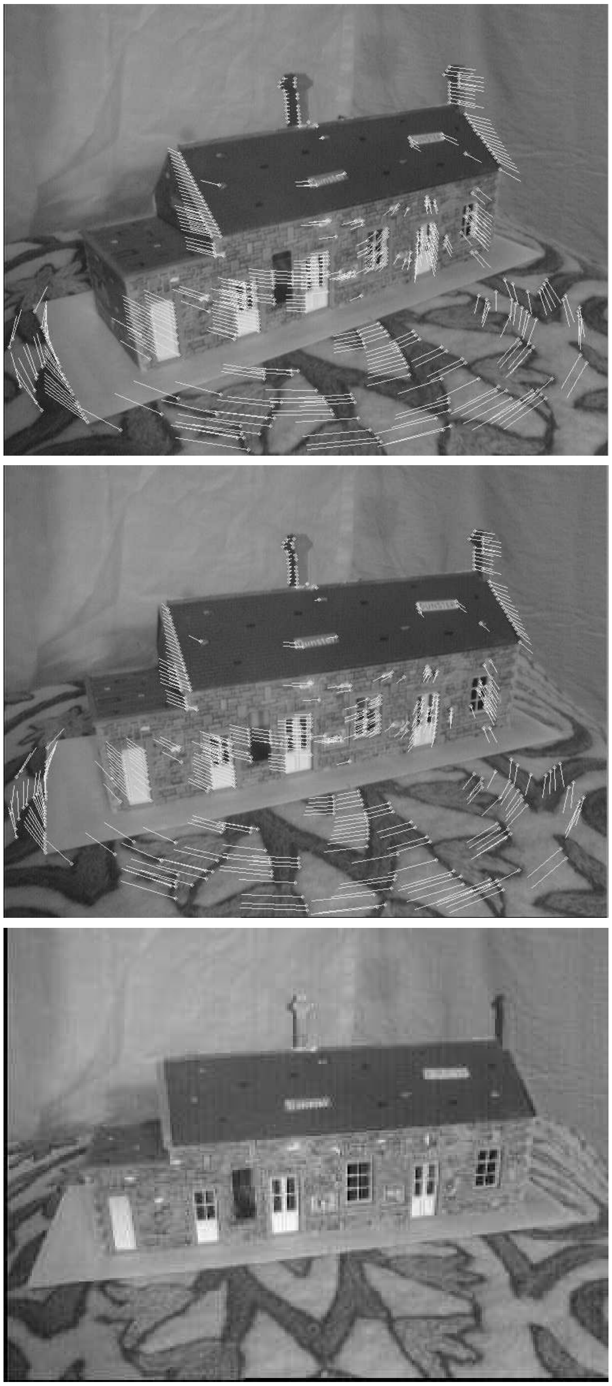

Fig. 3. Matched points with proposed method

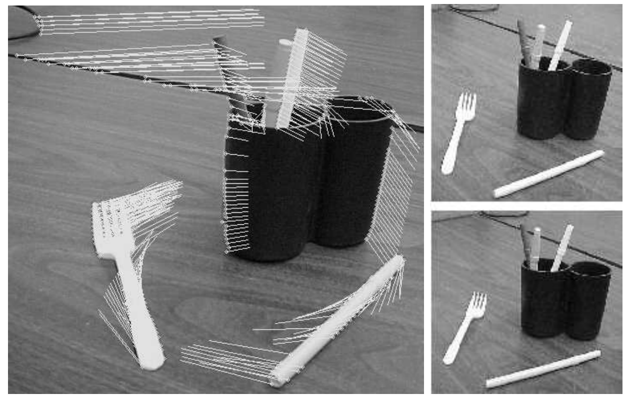

Fig. 4. Matched points with proposed method

8. REFERENCES

[1] A. Baumberg, Reliable Feature Matching Across Widely Seperated Views, *CVPR*, pp. 774-781, 2000

[2] C. Harris, M. Stephens, A Combined Corner and Edge Detector, *Alvey Vision Conf.*, 1988.

[3] P. Montesinos, V. Gouet, R. Deriche, D. Pelé, Matching color uncalibrated images using differential invariants, *Image and Vision Computing*, 2000.

[4] G. Roth, A. Whitehead, Using Projective vision to find Camera Positions in an Image Sequence, *Proc. of Vision Interface*, pp. 225-232, 2000.

[5] C. Schmid, R. Mohr, Local Grayvalue Invariants for Image Retrieval, *PAMI*, vol. 19, pp. 530-535, 1997.

[6] C. Schmid, R. Mohr, C. Bauckhage, Comparing and Evaluating Interest Points, *ICCV*, pp. 230-235, 1998.

[7] P. Torr, A. Zisserman, Robust Computation and Parameterization of Multiple View Relations, *ICCV*, pp. 727-732, 1998.

[8] T. Tuytelaars, L. Van Gool, L. D'haene, R. Koch, Matching of Affinely Invariant Regions for Visual Servoing, *ICRA*, pp. 1601-1606, 1999.

[9] E. Vincent, R. Laganire, Matching Feature Points for Telerobotics, *HAVE*, pp. 13-18, 2002.

[10] Z. Zhang, R. Deriche, O. Faugeras, Q.-T. Luong, A Robust Technique for Matching Two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geometry, *Artificial Intelligence*, 1995.