Grid-based Motion Statistics for Feature Correspondence

Hongzhi Liu

May 5, 2018

1 Feature Correspondence

There are always some unique pixels in an image. These points can be regarded as the feature of an image which are called feature points. Image feature correspondence based on the feature points is very important in the field of computer vision.

The concept of feature points has been widely used in the field of computer vision. The principle of this concept is to select some feature points from the image and analyze the image locally rather than to observe the whole image. These points are different and stable. Besides, they can be accurately located. Feature matching is the basic input of many computer vision algorithms. Furthermore, its speed and accuracy as well as robustness maintain a remarkable role in some relevant researches.

2 Grid-based Motion Statistics

Nowadays, there is a wide performance gap between slow feature matchers and the much faster real-time solutions. A number of techniques have focused on separating true and false matches using match distribution constraints, but their formulations result in complex smoothness constraints, which are difficult to understand and expensive to minimize. Besides, incorporating smoothness constraints into feature matching enables ultra-robust matching. However, such formulations are both complex and slow, making them unsuitable for video applications.

Professor Bian presents a simple means of encapsulating motion smoothness as the statistical likelihood of a certain number of matches in a region which is called *GMS* (Grid-based Motion Statistics). They show *GMS* can rapidly and reliably differentiate true and false matches, enabling high quality correspondence in Figure. 1. The cost of scoring each feature match's neighborhood is *O(N)*, where N is the number of image features. This conflicts with their desire to use as many features as possible. The team address it with a grid approximation that divides an image into $G = 20 \times 20$ nonoverlapping cells [1]. Scores of potential cell-pairs are computed only once.

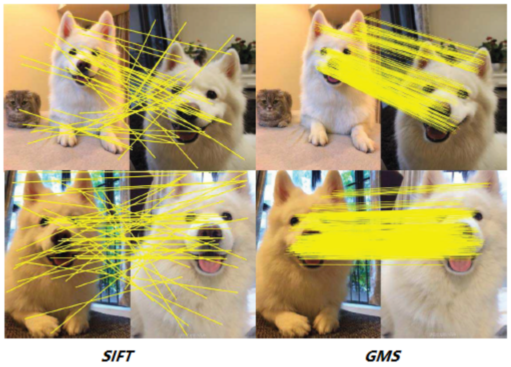

Figure 1: The highly respected SIFT [2] descriptor has difficulty on this scene because the dog's fur movement adds noise to local descriptors. Although they use weaker ORB descriptors, the GMS solution can leverage feature numbers to improve quality while maintaining real-time performance.

Grouping match neighborhoods (cell-pairs) for robustness. They give a more generalized score as equation (1) shows.

$$S_i = \sum_{k=1}^{K} |X_{a^k b^k}| - 1 \tag{1}$$

where K is the number of disjoint regions which match i predicts move together. They group cell-pairs using a smooth lateral motion assumption. Thus from equation (1), the score S_{ij} for cell-pair $\{i, j\}$ is shown in equation (2).

$$S_{ij} = \sum_{k=1}^{K=9} \left| X_{i^k j^k} \right| \tag{2}$$

where $\left| X_{i^k j^k} \right|$ is the number of matches between cells $\{i^k, j^k\}$ shown in Figure. 2.

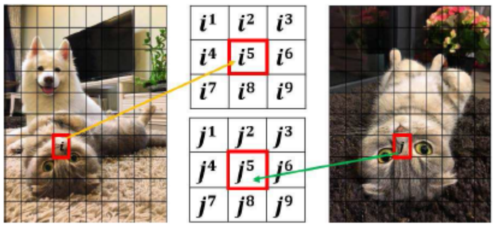

Figure 2: 9 regions around cells $\{i, j\}$ used in score evaluation.

Thresholding S_{ij} to divide cell-pairs into true and false sets $\{T, F\}$. In practice, m_f is small by design, while α is very large to ensure confident rejection of the large number of wrong cell-pair. The results in a single parameter thresholding function as shown in equations (3):

$$cell-pair \in \{i, j\} = \begin{cases} T, & \text{if} \quad S_{ij} > t_i = \alpha\sqrt{n_i} \\ F, & \text{otherwise} \end{cases} \tag{3}$$

where $\alpha = 6$ is experimentally determined and n_i is the average (of the 9 grid-cells in Figure. 2) number of features present in a single grid-cell.

3 Datasets and Evaluation

For evaluation, Professor Bian selects four datasets, TUM [5], Strecha [4], VGG [3] and Cabinet [5], described in Table. 1.

Professor Bian's team propose *GMS*, a statistical formulation for partitioning of true and false matches based on the number of neighboring matches. While this constraint has been implicitly employed by other techniques, their more principled approach enables development of simpler, faster algorithms with nearly equivalent performance.

Table 1: Dataset details. Strecha and VGG are standard benchmarks. TUM and Cabinet dataset are VGA resolution videos.

Dataset	TUM	Strecha	VGG	Cabinet
Full name	RGB-D SLAM Dataset	Dense Multiview Stereo Dataset	Affine Covariant Regions Datasets	A subset of TUM dataset
Image pairs	3,141	500	40	578
Ground truth	Camera pose	Camera pose	Homography	Camera pose
Description	Including all image	Well-textured images	Viewpoint change	Low-texture images

References

[1] Jiawang Bian, Wen Yan Lin, Yasuyuki Matsushita, Sai Kit Yeung, Tan Dat Nguyen, and Ming Ming Cheng. Gms: Grid-based motion statistics for fast, ultra-robust feature correspondence. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2828–2837, 2017.

[2] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[3] K Mikolajczyk, T Tuytelaars, C Schmid, A Zisserman, J Matas, F Schaffalitzky, T Kadir, and L. Van Gool. A comparison of affine region detectors. *International Journal of Computer Vision*, 65(1-2):43–72, 2005.

[4] C. Strecha, W. Von Hansen, L. Van Gool, P. Fua, and U. Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *Computer Vision and Pattern*

Recognition, 2008. CVPR 2008. IEEE Conference on, pages 1–8, 2008.

[5] Jrgen Sturm, Nikolas Engelhard, Felix Endres, Wolfram Burgard, and Daniel Cremers. A benchmark for the evaluation of rgb-d slam systems. In *Ieee/rsj International Conference on Intelligent Robots and Systems*, pages 573–580, 2012.