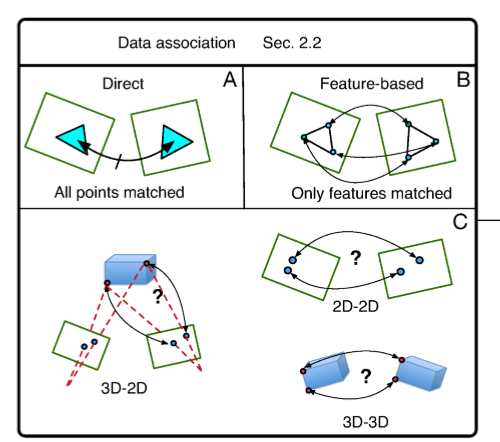
uses previous poses to guess poses for new frames

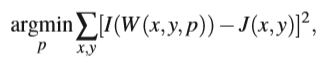


Data Association Choices – the data association method can be **direct, feature-based**, or a **hybrid** of the two. Direct methods use every pixel, whereas semi-dense methods only use pixels at which the gradient of image brightness is signiﬁcant.

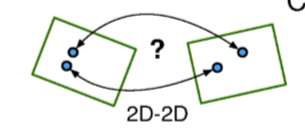
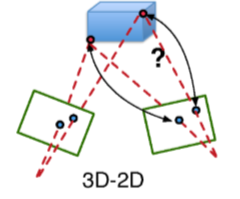
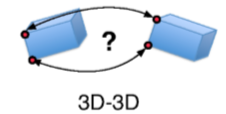
The basic underlying principle for all direct methods is known as the brightness consistency constraint and is best described as:

J(x,y,t) = I(x+u(x,y),y+v(x,y),t+1), (1)

where x and y are pixel coordinates; u and v denote displacement functions of the pixel (x,y) between two images I and J of the same scene taken at time t and t +1 respectively. The brightness consistency constraint is based on the assumption that a point from the world’s surface, observed in image I, will have the same intensity when observed in a subsequent image J. To render equation (1) solvable, [7] suggested, in what they referred to as Forward Additive Image Alignment (FAIA), to replace all the individual pixel displacements u and v by a single general motion model, in which the number of parameters is dependent on the implied type of motion. FAIA iteratively minimizes the squared pixel-intensity difference between the two images over the transformation parameters p:

where W(.,., p) is a warping transform that encodes the relationship relating the two images and p corresponds to the parameters of the transform. Equation (2) is non-linear and requires an iterative non-linear optimization process, with a computational complexity of O(n2N +n3) per iteration, where n is the number of parameters in p and N is the number of pixels in the image. Since 1981, other variants of the FAIA were suggested such as FCIA (Forward Compositional Image Alignment), ICIA (Inverse Compositional Image Alignment) and IAIA (Inverse Additive Image Alignment) each with different computational complexities. A detailed comparison between these variations can be found in [8].

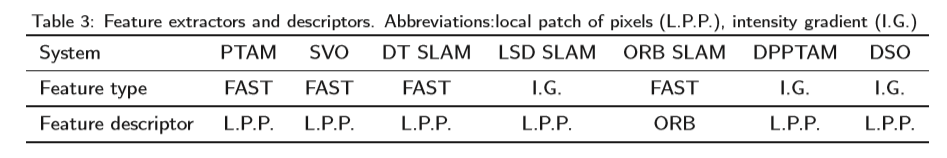
Data Association Types –

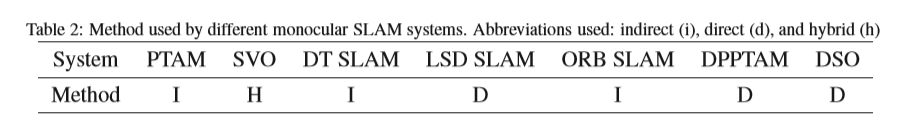
* 2D-2D
* 3D-2D
* 3D-3D

Feature-based methods were introduced to reduce the computational complexity of processing each pixel; this is done by matching only salient image locations, referred to as features, or keypoints. An example of feature-based matching is shown in Fig. 3-B. A descriptor is associated to each feature, which is used to provide a quantitative measure of similarity to other keypoints. On one hand, features are expected to be distinctive, invariant to viewpoint and illumination changes, as well as resilient to blur and noise; on the other hand, it is desirable for feature extractors to be computationally efﬁcient and fast. Unfortunately, such objectives are hard to achieve at the same time causing a trade-off between computational speed and feature quality. The computer vision community has developed, over decades of research, many different feature extractors and descriptors, each exhibiting varying performances in terms of ro

tation and scale invariance, as well as speed [9]. The selection of an appropriate feature detector depends on the platform’s computational power and the type of environment. Feature detector examplesinclude the Hessian corner detector [10], Harris detector [11], Shi-Tomasi corners [12], Laplacian of Gaussian detector [13], MSER [14], Difference of Gaussian [15] and the accelerated segment test family of detectors (FAST, AGAST, OAST) [16].

Feature descriptors include, and are not limited to, BRIEF [17], BRISK [18], SURF [19], SIFT [20], HoG [21], FREAK [22], ORB [23], and a low level local patch of pixels. Further information regarding feature extractors and descriptors is outside the scope of this work, but the reader can refer to [24], [25], [26], or [27] for comparisons.





**LSD SLAM:**

* Direct – extracts and makes use of all pixels that have a photometric gradient.
* Novel approach – performs direct, scale-drift aware alignment on sim(3), which is used to align two differently scaled keyframes.

DSO argues that using all the pixels with a photometric gradient introduces redundancy, and requires a regularization step: it subsamples the image by dividing the image into blocks, keeping a fixed number of pixels with the highest gradient in each block.

**Issues with Direct Methods:**

Direct methods exploit all information available in the image and are therefore more robust than feature-based methods in regions with poor texture and blur. Nevertheless, direct methods are susceptible to scene illumination changes, due to the violation of the underlying brightness consistency assumption (eq. (1) (J(x,y,t) = I(x+u(x,y),y+v(x,y),t+1))) . 🡺 In an effort to gain resilience against this mode of failure, the recently released DSO models the image formation process, and attempts to incorporate the scene irradiance in the energy functional, at the expense of adding a calibrated image formation model which is used to correct the images at a preprocessing step. The model is estimated through an additional ofﬂine calibration process described in [89].

During the non-linear optimization process, eq.(2 (argmin p

∑ x,y [I(W(x,y, p))− J(x,y)]2,)), is linearized through a ﬁrst order Taylor expansion. While the linearization is valid when the parameters of the warping transform tends to zero, higher order terms becomes dominant and the linearization becomes invalid for large transforms. Therefore, a second disadvantage of direct methods is the assumption of small motions between the images (typically not more than 1 pixel) 🡺 Direct monocular SLAM systems employ a pyramidal implementation, where the image alignment process takes place sequentially from the highest pyramid level to the lowest, using the results of every level as a prior to the next level. They also suggest the usage of high fame rate cameras to alleviate this issue; some systems employ an efﬁcient second order minimization (ESM [75]) to estimate a rotation prior that helps increase the convergence radius. Despite these efforts, the tolerated baseline for data association in direct methods is considerably smaller than the tolerated baseline in feature-based methods.

Another disadvantage of direct methods is that the calculation of the photometric error at every pixel is computationally intensive; therefore, real-time monocular SLAM applications of direct methods, until recently, were not considered feasible. However, with the recent advancements in parallelized processing and with the introduction of semi-dense inverse depth ﬁltering, it became possible to integrate direct methods into KSLAM solutions [63, 90, 66].