論 文 の 内 容 の 要 旨

論文題目　Discrete Inference Approaches to
Image Segmentation and Dense Correspondence
（画像領域分割と対応点推定問題への離散最適化アプローチ）

氏　　名　谷合　竜典

We consider discrete inference approaches to image segmentation and dense correspondence. The two problems cover diverse tasks such as image segmentation, binarization, cosegmentation, motion segmentation, binocular stereo vision, optical flow and general dense correspondence, which are addressed sorely or jointly in this work as energy minimization problems on Markov random fields (MRFs). Discrete inference approaches are employed to effectively optimize inherently discrete functions or highly non-convex continuous functions. The contributions of this work are two folds: proposal of novel joint frameworks of image segmentation and dense correspondence problems, and development of new inference techniques for sole or joint tasks. Specifically, we comprehensively address three challenges of discrete inference, that is, label space size, higher-order energy, and non-submodular energy, which are posed in various forms in the following tasks that we tackle.

First, we study inference problems on non-submodular and higher-order MRFs that have binary variables. Such problems naturally appear in low-level computer vision tasks such as image segmentation and binarization of gray images. They are also imposed as subproblems in estimation of more general multi-valued or continuous-valued variables. For such fundamental inference problems, we develop a new theoretical insight into several existing optimization methods and propose a new method by unifying them. The proposed method has a mechanism to better avoid bad local minimums of non-submodular functions, and is thus more robust to initializations compared to existing methods. The proposed method was evaluated on image segmentation and binarization tasks and was shown to outperform state-of-the-art methods.

Second, we propose an efficient and accurate binocular stereo matching method, whose model and inference both favor piecewise planar surfaces. We formulate the stereo problem by a model of per-pixel local 3D surface planes with piecewise planar smoothness regularization, which forms a pairwise MRF with a continuous 3D label space. In order to efficiently infer this rich model, we propose a new inference technique that extends the well-known expansion move algorithm by incorporating the spatial propagation and randomization search mechanisms of PatchMatch inference. Unlike conventional fusion-based approaches, the proposed method does

not require solution proposals and also produces submodular energies that are optimally minimized by graph cuts during the inference. The computations can be easily accelerated by parallelization and using fast cost-map filtering. The proposed method achieved the state-of-the-art performance on the Middlebury stereo benchmark among more than 160 stereo algorithms.

Third, we propose a unified framework of general dense correspondence and cosegmentation for two images, where common "foreground" regions in the two images are segmented and aligned to each other. Our method is formulated using a hierarchical MRF model with joint labels of segmentation and correspondence. The correspondence field is parameterized using similarity transformations (4-DOF) assigned on superpixels. The hierarchy is used to evaluate correspondence across various coarseness of superpixels, which brings high robustness when aligning objects with different appearances. Unlike prior hierarchical methods which assume that the structure is given, we dynamically recover the structure along with the correspondence and segmentation labeling. This joint inference is performed in an energy minimization framework using iterated graph cuts. The proposed method was quantitatively evaluated on a new dataset and it outperformed state-of-the-art methods designed specifically for either cosegmentation or correspondence estimation.

Finally, we propose a fast scene flow method for stereo image sequences that simultaneously recovers motion segmentation of moving objects as well as camera ego-motion. This framework unifies four tasks —stereo, optical flow, motion segmentation and visual odometry providing rich information of disparity, 2D flow and binary segmentation of moving objects at every pixel along with camera motion. The inference is carried out through a multi-staged pipeline where the solution to one task benefits others, leading to overall higher accuracy and efficiency. The proposed method was evaluated on the KITTI 2015 scene flow benchmark and was ranked third. Furthermore, our CPU implementation processed each frame in 2-3 seconds, which was 1-3 orders of magnitude faster than the top six methods that took 1-50 minutes per frame. Our method was also thoroughly evaluated on challenging Sintel sequences having fast camera and object motion, where our method consistently outperformed the method ranked second on the KITTI benchmark.