Journal of Visual Communication and Image Representation 55 (2018) 529–539

Contents lists available at ScienceDirect

Journal of Visual Communication and Image Representation

journal homepage: www.elsevier.com/locate/jvci

Efficient graph cut optimization for shape from focus☆

Christophe Ribal*, Nicolas Lermé, Sylvie Le Hégarat-Mascle

Université Paris-Sud, Laboratoire SATIE – UMR CNRS 8029, Rue Noetzlin, 91190 Gif-sur-Yvette, France

ARTICLE INFO

Keywords:
Shape from focus
Depth map estimation
Graph cuts
Multi-labels

ABSTRACT

Shape From Focus refers to the inverse problem of recovering the depth in every point of a scene from a set of differently focused 2D images. Recently, some authors stated it in the variational framework and solved it by minimizing a non-convex functional. However, the global optimality on the solution is not guaranteed and evaluations are often application-specific. To overcome these limits, we propose to globally and efficiently minimize a convex functional by decomposing it into a sequence of binary problems using graph cuts. To illustrate the genericity of such a decomposition-based approach, data-driven strategies are considered, allowing us to optimize (in terms of reconstruction error) the choice of the depth values for a given number of possible depths. We provide qualitative and quantitative evaluation on Middlebury datasets and we show that, according to classic statistics on error values, the proposed approach exhibits high performance and robustness against corrupted data.

1. Introduction

1.1. Context

Retrieving the depth of a scene from a collection of at least one image is a challenging inverse problem that is typically solved using shape-from-X approaches (where X denotes the cue to infer the shape, e.g. stereo, motion, shading, focus, defocus, etc) or a mixture of them. This topic gave rise to a huge amount of papers and still represents a great interest for researchers in the computer vision community. Indeed, it has numerous applications, especially in robotics, both for localization and environment analysis, in monitoring or video-surveillance either for security or for medical technical assistance, or in microscopy and chemistry [1].

More specifically, let us remind that stereovision relies on the disparities between matched pixels of an image pair [2], shape-from-shading exploits the variations of brightness of a single image [3,4] and shape-from-motion deduces depth from matched points of interest [5]. Shape-from-focus (SFF) [6] and shape-from-defocus (SFD) [7] represent alternatives approaches that share the idea of using the focus to estimate the 3D structure of a scene from differently focused images acquired by a monocular camera. Thus, an object appears focused only in a limited range (depth of field) and is progressively blurred as the camera moves away from this range. For both approaches, active and passive sensors exist, depending on whether or not a structured light composed of patterns is projected onto the scene to alleviate ambiguities. In this paper, we will focus on the passive device. In addition to the depth map, both approaches generally also provide an estimation of the all-in-focus image of the scene, i.e. the image obtained by selecting for each pixel, the intensity at which it appears the most focused, or sharp.

Now, SFF and SFD differ on one main point. SFD estimates the depth by measuring the relative blurriness between a reference image and the remaining ones. The blurring process needs to be explicitly modeled, a very few images are usually required and the approach can be applied to dynamic scenes. Similarly, [8,9] have chosen to solve the inverse problem by precisely modeling the defocusing process with the help of an all-in-focus image. This requires the knowledge of the parameters of the camera to compute the spatially varying point spread function (PSF). In these works, the authors iteratively minimize, using Split Bregman algorithm, a regularized energy computed from the distance between the observations and the approximated PSFs applied to the all-in-focus images.

SFF only assumes that there is an explicit relationship between the depth of a given pixel and the focal value at which it appears the most focused (or sharp). This implies the choice of an appropriate predefined operator for measuring the amount of sharpness, and a fairly large number of images to expect a good reconstruction quality of the scene. Therefore, SFF is mainly used to analyze static scenes.

In contrast to multi-cameras systems, SFF and SFD approaches allow

☆ This paper has been recommended for acceptance by Yong Jae Lee.

* Corresponding author.
E-mail address: christophe.ribal@u-psud.fr (C. Ribal).

https://doi.org/10.1016/j.jvcir.2018.06.029
Received 9 February 2018; Received in revised form 3 May 2018; Accepted 30 June 2018
Available online 11 July 2018

for a more compact size of the electronic system, decrease its cost and avoid to deal with matching ambiguities. The topic is still of interest as demonstrated by recent works, e.g. [8,10,9,11], including machine learning with convolutional neural networks [11].

1.2. Related work

As previously explained, solving the SFF problem implies the choice of an appropriate sharpness operator for selecting the focus maximizing the pixel sharpness. First among many, Nayar [6] introduces a sharpness operator named Summed Modified LAPlacian (SMLAP) based on second derivatives. Then, we refer the reader to the study [12] that compares a wide variety of sharpness operators in a comprehensive way.

The idea of early approaches (such as [6]) is to compute a sharpness profile over the focus values and take the argument of the maximum of this profile for every pixel. However, whatever the used sharpness operator, an estimation using raw profile is prone to errors in presence of degraded or noisy data so that different filters adapted to the sharpness profile have been proposed. In [6], a Gaussian interpolation is performed around the maximum detected on the raw profile. As an alternative to Gaussian interpolation, [13] proposed to interpolate the sharpness profile by a low-order polynomial. This idea has been then followed in [10], in which an eight-order polynomial is used.

Whatever the sharpness operator and the interpolation method used, blind techniques (i.e. that consider pixels independently of their neighbors) do not generally allow for accurate recovering the 3D geometry of a whole scene. Indeed, the sharpness operator relies on object borders that produce sharp edges on which reliable and precise depth values may be deduced. In the absence of such elements or of texture, the maximum of sharpness location tends to produce unreliable results. Ambiguities are especially present in textureless, underexposed or overexposed regions. To cope with these problems, some authors [14] proposed to reject the sharpness values being under a threshold, resulting in a globally more reliable, but sparse depth map.

Since the measurements from sharpness operator do not necessarily determine the depth uniquely, the SFF is an ill-posed problem. While formulating this kind of problem in the variational framework is a standard way to tackle it, surprisingly, only very few papers [15,17,10] did it. Mathematically, this amounts to the definition of a functional that embeds a data fidelity term and a smoothness (or regularization) term and that has to be (efficiently) minimized.

In [10], the variational formulation uses the negative interpolated contrast measure from Modified LAPlacian (MLAP, i.e. SMLAP restricted to a single pixel) as data fidelity term. As a result, this term is a non-convex but smooth continuous function. The regularization term used is the discrete isotropic Total Variation (TV), discontinuity-preserving, non-smooth but convex. To minimize the resulting non-convex functional, the data term is linearized and an iterative algorithm, namely Alternating Direction Method of Multipliers (ADMM) is applied. According to the authors, this algorithm provably converges toward a critical point of the functional but no optimality guarantees are mentioned about the solution. Although the proposed algorithm seems to give good results and exhibit good convergence properties, it has been actually evaluated only qualitatively and on few real images.

The work of [15] also uses the sharpness operator MLAP. The data fidelity term is the truncated quadratic difference between the maximum value of sharpness and the tested sharpness. This term is therefore non-convex. The smoothness term is a truncated L^2 norm (then also non-convex) that is discontinuity-preserving. The truncation depends on whether a significant texture is present or not. The algorithm used for the minimization of the resulting non-convex functional is the α-expansion based on graph cuts [16]. Interesting results are obtained but the approach is prone to get easily trapped in local minima of the energy and in [15], the evaluation is limited to application-specific images (optical microscopes).

Table 1
Functional properties between the proposed approach, [15,10].

	Our	[15]	[10]
Data term	Convex	Non-convex	Non-convex
Regularization term	Convex	Non-convex	Convex
Functional	Convex	Non-convex	Non-convex
Optimization method	Graph cut	Graph cut	ADMM
Optimality	Globally optimal	Within a known factor of the global minimum [16]	No guaranty of optimality

1.3. Outline of the proposed approach

In this work, we explore a new way to solve the SFF problem by directly minimizing, for a given depth resolution, a *convex* functional. The advantage of such a choice is twofold: (i) The optimality about the solution is easier to guarantee and (ii) the convexity property can be exploited to use fast minimization procedures. Functional properties of the aforementioned approaches against ours are summarized in the Table 1.

Our choice focuses on graph cuts because of their well-founded theoretical background [18] and the existence of a fast maximum-flow/minimum-cut algorithm [19]. While [20] has optimality guarantees for convex priors, the graph construction requires a lot of computational resources (in terms of time and memory). Alternatives, like the α-expansion [16], allow for minimizing the functional iteratively by solving sequentially binary problems until convergence, but without any guaranty relatively to the number of iterations required.

Thanks to a discretization step, the functional can nevertheless be exactly minimized when the data fidelity term is convex, by mapping the original problem to a deterministic number of independent binary problems (each one solved using graph cuts) [21]. Each subproblem boils down to choosing a split value along the depth dimension and labeling the depth map accordingly. Given a number of binary problems, the dyadic strategy is an usual efficient way to select these split values, but a data-driven splitting strategy allows for lowering the reconstruction error, especially when the fixed number of discrete depth values is low. Another beneficial effect is to balance the sizes of the subproblems, thus reducing the complexity of the divide-and-conquer approach.

Fig. 1 gives the outline of our approach. Our optimization algorithm is based on graph cuts (bottom right rectangular box on Fig. 1). Besides data images and regularization parameter λ, it takes as an input the tree of *split values*, i.e. the values that define hierarchically the subproblems.

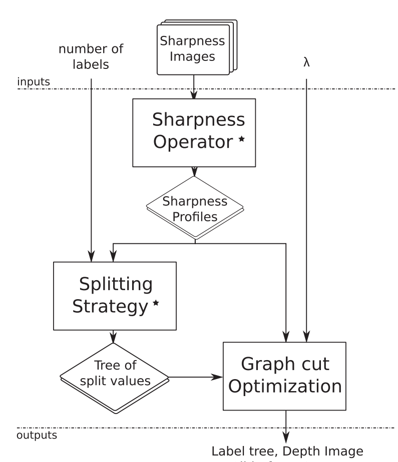

Fig. 1. Flowchart representing the approach chosen to solve our problem. In this paper, different implementations are explored for starred boxes.