Medical Image Analysis 33 (2016) 102–106

Contents lists available at ScienceDirect

Medical Image Analysis

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

Editorial

(Hyper)-graphical models in biomedical image analysis

Nikos Paragios[a,*], Enzo Ferrante[c], Ben Glocker[c], Nikos Komodakis[b], Sarah Parisot[c], Evangelia I. Zacharaki[a]

[a] CentraleSupelec, Inria, Université Paris-Saclay, France
[b] Ecole des Ponts ParisTech, Université Paris-Est, France
[c] Imperial College of London, London, UK

ARTICLE INFO

Article history:
Received 20 April 2016
Revised 16 June 2016
Accepted 16 June 2016
Available online 23 June 2016

Keywords:
(Hyper)graphs
Random fields
Message passing
Graph cuts
Linear programming
Image segmentation
Shape & volume registration

ABSTRACT

Computational vision, visual computing and biomedical image analysis have made tremendous progress over the past two decades. This is mostly due the development of efficient learning and inference algorithms which allow better and richer modeling of image and visual understanding tasks. Hyper-graph representations are among the most prominent tools to address such perception through the casting of perception as a graph optimization problem. In this paper, we briefly introduce the importance of such representations, discuss their strength and limitations, provide appropriate strategies for their inference and present their application to address a variety of problems in biomedical image analysis.

1. Introduction

The analysis of medical images aims at extraction of hidden, clinically relevant information, and has a wide range of applications including the detection or automatic outlining of internal structures (image segmentation), the summarization of knowledge from group populations (spatial normalization), or the association of imaging profile with the clinical state of the patient. Inverse modeling is often the paradigm used to interpret medical images. In simple words given an interpretation objective, the first step consists on describing the space of solutions through a parametric mathematical model. The parameters of this model are then somehow associated with the measurements through an objective specific cost function that also imposes some smoothness constraints to make the problem mathematically and computationally tractable. Optimal solution to the problem consists in finding the set of parameters able to produce the lowest cost of the energy function and is often determined through an approximate inference algorithm since exact solution is frequently not attainable.

Numerous mathematical paradigms have been considered in the field of medical imaging to implement the aforementioned strategy. Mathematical richness, computational complexity, inference mechanisms and modularity and scalability are often their classification criteria. Ideally one would like to adopt models of limited complexity (it facilitates inference) but able to encode rich priors, as well as cost functions that can be exactly minimized (optimal solution) with reasonable computational complexity. Graphical models (Koller and Friedman, 2009) is an elegant, powerful and computationally efficient method to such an objective. In the recent years, the field has witnessed an enormous progress due to the development of efficient optimization/inference methods coupled with machine learning algorithms and the availability of large scale training data. While probabilistic graphical models have a variety of useful variants, here we will focus on a Markov Random Fields (MRF) formulation, where inference is often expressed as a (undirected) graph optimization problem acting on a predefined graph structure (fixed number of nodes and connectivity) associated with a discrete number of variables.

A wide variety of tasks in medical image analysis can be formulated as discrete labeling problems. In very simple terms, a discrete optimization problem can be stated as follows: we are given a discrete set of variables $\mathcal{V}$, all of which are vertices in a graph $\mathcal{G}$. The edges of this graph (denoted by $\mathcal{E}$) encode the variables' relationships. We are also given as input a discrete set of labels $\mathcal{L}$. We must then assign one label from $\mathcal{L}$ to each variable in $\mathcal{V}$. However, each time we choose to assign a label, say, x_{p_1} to a variable p_1, we are forced to pay a price according to the so-called *singleton* potential function $g_p(x_p)$, while each time we choose to assign a pair of

* Corresponding author. Fax: +33 1 41 13 1006.
E-mail address: nikos.paragios@centralesupelec.fr (N. Paragios).

http://dx.doi.org/10.1016/j.media.2016.06.028
1361-8415/

labels, say, x_{p_1} and x_{p_2} to two interrelated variables p_1 and p_2 (two nodes that are connected by an edge in the graph $\mathcal{G}$), we are also forced to pay another price, which is now determined by the so called *pairwise* potential function $f_{p_1p_2}(x_{p_1}, x_{p_2})$. Both the singleton and pairwise potential functions are problem specific and are thus assumed to be provided as input. For example, if the aim is image segmentation, in order to assign a class label to every voxel of the image, an appearance model is learned (from training examples) for every class and encoded in the form of a probability distribution $\pi(x)$. This probability distribution can be naturally incorporated in the MRF model by setting the unary potentials of the segmentation grid for every label to the negative log-probability of the respective class:

$$g_p(x_p) = -log(\pi(x_p)).$$

Our goal is then to choose a labeling which will allow us to pay the smallest total price. In other words, based on what we have mentioned above, we want to choose a labeling that minimizes the sum of all the MRF potentials, or equivalently the MRF energy. This amounts to solving the following optimization problem:

$$\arg\min_{\{x_p\}} \mathcal{P}(g, f) = \sum_{p\in\mathcal{V}} g_p(x_p) + \sum_{(p_1,p_2)\in\mathcal{E}} f_{p_1p_2}(x_{p_1}, x_{p_2}). \quad (1)$$

The use of such a model can describe a number of challenging problems in medical image analysis. However these simplistic models can only account for simple interactions between variables, a rather constrained scenario for high-level medical imaging perception tasks. One can augment the expressive power of this model through higher order interactions between variables, or a number of cliques $\{C_i, i \in [1, n] = \{\{p_{i1}, \cdots, p_{i|C_i|}\}\}$ of order $|C_i|$ that will augment the definition of $\mathcal{V}$ and will introduce hyper-vertices:

$$\arg\min_{\{x_p\}} \mathcal{P}(g, f) = \sum_{p\in\mathcal{V}} g_p(x_p) + \sum_{(p_1,p_2)\in\mathcal{E}} f_{p_1p_2}(x_{p_1}, x_{p_2}) + \sum_{C_i\in\mathcal{E}} f_{p_1\cdots p_n}(x_{p_{i1}}, \cdots, p_{x_{i|C_i|}}). \quad (2)$$

where $f_{p_1\cdots p_n}$ is the price to pay for associating the labels $(x_{p_{i1}}, \ldots, p_{x_{i|C_i|}})$ to the nodes $(p_1 \cdots p_{i|C_i|})$. Parameter inference, addressed by minimizing the problem above, is the most critical aspect in computational medicine and efficient optimization algorithms are to be evaluated both in terms of computational complexity as well as of inference performance. State of the art methods include deterministic and non-deterministic annealing, genetic algorithms, max-flow/min-cut techniques and relaxation. These methods offer certain strengths while exhibiting certain limitations, mostly related to the amount of interactions which can be tolerated among neighborhood nodes. In the area of medical imaging where domain knowledge is quite strong, one would expect that such interactions should be enforced at the largest scale possible.

The reminder of this paper reviews briefly our contributions in the field. Section 2 presents the work done in the area of inference algorithms while Section 3 discusses their use to address the fundamental problems of biomedical image analysis, while the last section concludes the paper and presents perspectives and future directions of our work.

2. Inference on graphical models

Inference over graphs has been a well studied problem in a number of fields such as networks, operational research, computational biology and computer vision. Graph-based representations were introduced in computer vision at mid-eighties (Geman and Geman, 1984) through Markov Random Fields as a novel mathematical modeling framework constrained though from the lack of efficient inference methods as well as processing power - and became again popular during the past two decades thanks to the development of efficient optimization algorithms (Kappes et al., 2015; Wang et al., 2013).

On one hand, local iterative optimization methods or annealing-like methods were the first attempts to perform inference over graphs in computer vision and image processing. Computational efficiency was the main strength of local iterative methods with bottleneck being their inability to converge to a good optimum. Annealing methods - despite their theoretical guarantees with respect to the attained solution - were computationally inefficient and practically unusable. The principle of max-flow/min-cut theorem was first introduced in vision at the late eighties (Greig et al., 1989) and then re-introduced efficiently (Boykov et al., 1998) along with the re-discovery of message passing methods (Pearl, 1982) despite the absence of theoretical guarantees - such as belief propagation networks (Yedidia et al., 2000). Such developments had as a direct consequence the establishment of graphical models as one of the main stream computational vision mathematical formalism over the past decade. In the recent years, a number of extremely efficient optimization methods were re-introduced (e.g. Tree-reweighted Message Passing (Kolmogorov, 2006), the Fast Primal-Dual Method (Komodakis et al., 2008) or the Extended Roof Duality (Rother et al., 2007)) resulting on a great variety of optimization algorithms addressing the expectations of the field both in terms of optimality properties of the attained solution as well as in terms of computational complexity for low-rank (pair-wise) graphical models.

On the other hand, the case of higher order models with arbitrary interactions between nodes (both in terms of connectivity as well as in terms of potentials) is far from being considered that has reached a mature stage. Significant progress has been made when considering higher order interactions of simple nature such as generalized Potts model (Kohli et al., 2009) or higher order models of limited complexity through their mapping to a pair-wise one and then use existing optimization methods to perform inference (Ishikawa, 2011). In the case of real-valued variables, an alternative consists of using message passing methods like belief propagation networks (Potetz and Lee, 2008) that have linear complexity with respect to the order of clique. Dual decomposition (Komodakis et al., 2011) is another alternative that provides more freedom with respect to the class of potentials that could be handled while maintaining a competitive advantage in terms of computational complexity. In Fix et al. (2015) a more efficient variant for mapping higher order problems to pair-wise ones using principles introduced in Komodakis et al. (2008) was presented with better convergence and optimality guarantees. Furthermore, in Arora et al. (2015) an efficient method to infer solutions for higher order graphical models with submodular potentials was presented, while in Osokin and Vetrov (2015) a submodular relaxation approach was proposed for pair-wise and higher order graphical models inference. More recently, Shekhovtsov (2016) tackled higher order inference for graphical models which can be expressed as the minimization of partially separable function of discrete variables, while in Khandelwal et al. (2016) the principle of active constraints adaptively learnt over multiple iterations was adopted. Further aspects on inference, and specifically on the most commonly used optimization principles in the context of graphical models, are discussed in Komodakis et al. (2016).

3. Discrete biomedical image analysis

Graph-based representations have attracted the interest of the biomedical image analysis community immediately after their reappearance in the field of vision. Discrete labeling problems like semantic image segmentation were the first to be considered. Let