

Robust Graph-Cut Scene Segmentation and Reconstruction for Free-Viewpoint Video of Complex Dynamic Scenes – Supplementary Material

Jean-Yves Guillemaut, Joe Kilner and Adrian Hilton

Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, GU2 7XH, UK

{j.guillemaut, j.kilner, a.hilton}@surrey.ac.uk

Abstract

In [3], we proposed a technique for robust joint multi-layer segmentation and reconstruction of large scale outdoor scenes from multiple loosely calibrated moving cameras separated by a wide-baseline. The technique combines multiple visual cues such as dense photo-consistency scores, sparse affine covariant feature constraints, probabilistic colour models of layers, contrast and smoothness priors into an energy function which must be minimised. The resulting energy function is non-convex and has a large number of variables. Global minimisation of this type of energy function is known to be NP-hard. In this technical report, we demonstrate that the defined energy function satisfies the regularity constraint required for graph-cut optimisation via the expansion move algorithm. This allows computation of a strong local optimum which is guaranteed to be within a known factor of the global optimum.

This technical report is structured as follows. We start by rewriting the energy function in a form suitable for the proof of regularity. Then we give a brief description of graph-cut optimisation via expansion move algorithm. Finally we demonstrate that the energy function satisfies the regularity condition required for minimisation using this framework.

1. Energy function formulation

The energy function is defined in Section 3 of [3]. In this technical report, we use the same notation as in [3]. Whenever the definition of a variable has been omitted, the reader is referred to [3]. The energy function that we seek to optimise can be written in its generic form as

$$E(l, d) = \sum_{\mathbf{p} \in \mathcal{P}} D_{\mathbf{p}}(l_{\mathbf{p}}, d_{\mathbf{p}}) + \sum_{(\mathbf{p}, \mathbf{q}) \in \mathcal{N}} V_{\mathbf{p}, \mathbf{q}}(l_{\mathbf{p}}, d_{\mathbf{p}}, l_{\mathbf{q}}, d_{\mathbf{q}}). \quad (1)$$

$D_{\mathbf{p}}(l_{\mathbf{p}}, d_{\mathbf{p}})$ is a unary term which measures the cost of assigning a label $(l_{\mathbf{p}}, d_{\mathbf{p}})$ to pixel \mathbf{p} based on the observed data. It is defined as

$$D_{\mathbf{p}}(l_{\mathbf{p}}, d_{\mathbf{p}}) = D'_{\mathbf{p}}(l_{\mathbf{p}}, d_{\mathbf{p}}) + D''_{\mathbf{p}}(l_{\mathbf{p}}, d_{\mathbf{p}}), \quad (2)$$

where

$$D'_{\mathbf{p}}(l_{\mathbf{p}}, d_{\mathbf{p}}) = -\lambda_{\text{colour}} \log P(\mathbf{I}_{\mathbf{p}} | l_{\mathbf{p}}), \quad (3)$$

and

$$D''_{\mathbf{p}}(l_{\mathbf{p}}, d_{\mathbf{p}}) = \lambda_{\text{match}} (e_{\text{dense}}(\mathbf{p}, d_{\mathbf{p}}) + e_{\text{sparse}}(\mathbf{p}, d_{\mathbf{p}})). \quad (4)$$

The term $D'_{\mathbf{p}}$ uses learnt colour models for each layer to encourage assignment of pixels to the layer following the most similar colour model, while $D''_{\mathbf{p}}$ encourages depth assignments to maximise a dense photo-consistency measure as well as some sparse affine covariant feature constraints.

$V_{\mathbf{p}, \mathbf{q}}(l_{\mathbf{p}}, d_{\mathbf{p}}, l_{\mathbf{q}}, d_{\mathbf{q}})$ is a pairwise term which measures the cost of assigning a pair of labels $(l_{\mathbf{p}}, d_{\mathbf{p}})$ and $(l_{\mathbf{q}}, d_{\mathbf{q}})$ to neighbouring pixels \mathbf{p} and \mathbf{q} based on some priors. It is defined as

$$V_{\mathbf{p}, \mathbf{q}}(l_{\mathbf{p}}, d_{\mathbf{p}}, l_{\mathbf{q}}, d_{\mathbf{q}}) = V'_{\mathbf{p}, \mathbf{q}}(l_{\mathbf{p}}, d_{\mathbf{p}}, l_{\mathbf{q}}, d_{\mathbf{q}}) + V''_{\mathbf{p}, \mathbf{q}}(l_{\mathbf{p}}, d_{\mathbf{p}}, l_{\mathbf{q}}, d_{\mathbf{q}}), \quad (5)$$

where

$$V'_{p,q}(l_p, d_p, l_q, d_q) = \lambda_{\text{contrast}} \begin{cases} 0 & \text{if } l_p = l_q, \\ \exp(-\beta C(\mathbf{I}_p, \mathbf{I}_q)) & \text{otherwise,} \end{cases} \quad (6)$$

and

$$V''_{p,q}(l_p, d_p, l_q, d_q) = \lambda_{\text{smooth}} \begin{cases} \min(|d_p - d_q|, d_{\max}) & \text{if } l_p = l_q \text{ and } d_p, d_q \neq \mathcal{U}, \\ 0 & \text{if } l_p = l_q \text{ and } d_p, d_q = \mathcal{U}, \\ d_{\max} & \text{otherwise.} \end{cases} \quad (7)$$

The term $V'_{p,q}$ encourages discontinuities to follow high contrast regions, while the term $V''_{p,q}$ encourages the depth labels to vary smoothly within each layer.

2. Graph-cut optimisation using the expansion move algorithm

In order to compute a solution to the joint segmentation and reconstruction problem, the previous energy function must be minimised. Because the energy function defined is non-convex and has a large number of variables, finding a global minimum is NP-hard. However, recent work on graph-cut optimisation in [2] has shown that it is possible to compute a strong local minimum which is guaranteed to be within a known factor of the global optimum by using the expansion move algorithm. The main idea of the algorithm is to cycle through the set of labels and perform an expansion move with respect to each label until the energy cannot be decreased. Each expansion move requires solving a binary optimisation problem for which a global solution is obtained using the min-cut/max-flow algorithm [1]. In [4], Kolmogorov and Zabih demonstrate that in order to be minimisable via graph-cuts, the energy function must satisfy a property called regularity (also called sub-modularity). In the next section, we demonstrate that the energy function considered satisfies this property.

3. Proof of regularity

According to [4], the energy is graph representable if and only if each pairwise term satisfies the inequality

$$V_{p,q}(l_p, d_p, l_q, d_q) + V_{p,q}(l_\alpha, d_\alpha, l_\alpha, d_\alpha) \leq V_{p,q}(l_p, d_p, l_\alpha, d_\alpha) + V_{p,q}(l_\alpha, d_\alpha, l_q, d_q). \quad (8)$$

Functions satisfying this inequality are called regular and can be efficiently optimised via graph-cuts. For the energy function considered, we have $V_{p,q}(l_\alpha, d_\alpha, l_\alpha, d_\alpha) = 0$, therefore the previous inequality simplifies to

$$V_{p,q}(l_p, d_p, l_q, d_q) \leq V_{p,q}(l_p, d_p, l_\alpha, d_\alpha) + V_{p,q}(l_\alpha, d_\alpha, l_q, d_q). \quad (9)$$

We show that both the contrast terms $V'_{p,q}$ and the smoothness terms $V''_{p,q}$ satisfy this inequality. This is a sufficient condition for the regularity constraint to be satisfied.

3.1. Contrast term

This term does not depend on depth labels. We distinguish the following cases which cover all possible label configurations.

Case 1: $l_p = l_q$ Then $V'_{p,q}(l_p, d_p, l_q, d_q) = 0$ and the regularity constraint is equivalent to

$$0 \leq V'_{p,q}(l_p, d_p, l_\alpha, d_\alpha) + V'_{p,q}(l_\alpha, d_\alpha, l_q, d_q), \quad (10)$$

which is always satisfied because $V'_{p,q}$ is by definition positive.

Case 2: $l_p \neq l_q, l_p \neq l_\alpha, l_q \neq l_\alpha$ Each term $V'_{p,q}$ is constant with respect to the labels and the regularity constraint is trivially satisfied.

Case 3: $l_p \neq l_q, l_p = l_\alpha, l_q \neq l_\alpha$ We have $V'_{p,q}(l_p, d_p, l_q, d_q) = V'_{p,q}(l_\alpha, d_\alpha, l_q, d_q)$ so the regularity constraint is equivalent to

$$0 \leq V'_{p,q}(l_p, d_p, l_\alpha, d_\alpha), \quad (11)$$

which is always satisfied because $V'_{p,q}$ is by definition positive.

Case 4: $l_p \neq l_q, l_p \neq l_\alpha, l_q = l_\alpha$ The proof is identical to the previous case after permutation of l_p and l_q .

3.2. Smoothness term

We distinguish the following cases which cover all possible label configurations.

Case 1: $l_p = l_q = l_\alpha$ We consider the following sub-cases.

Sub-case i: $d_p = d_q = \mathcal{U}$ We have $V''_{p,q}(l_p, d_p, l_q, d_q) = 0$ and the regularity constraint is equivalent to

$$0 \leq V''_{p,q}(l_p, d_p, l_\alpha, d_\alpha) + V''_{p,q}(l_\alpha, d_\alpha, l_q, d_q), \quad (12)$$

which always holds because by definition $V''_{p,q}$ is positive.

Sub-case ii: $d_p = \mathcal{U}, d_q \neq \mathcal{U}$ If $d_\alpha = \mathcal{U}$, we have $V''_{p,q}(l_p, d_p, l_q, d_q) = V''_{p,q}(l_\alpha, d_\alpha, l_q, d_q)$ and the regularity constraint is equivalent to

$$0 \leq V''_{p,q}(l_p, d_p, l_\alpha, d_\alpha). \quad (13)$$

If $d_\alpha \neq \mathcal{U}$, we have $V''_{p,q}(l_p, d_p, l_q, d_q) = V''_{p,q}(l_p, d_p, l_\alpha, d_\alpha)$ and the regularity constraint is equivalent to

$$0 \leq V''_{p,q}(l_\alpha, d_\alpha, l_q, d_q). \quad (14)$$

In both cases the inequality is satisfied because $V''_{p,q}$ is by definition positive.

Sub-case iii: $d_p \neq \mathcal{U}, d_q = \mathcal{U}$ The proof is identical to the previous case after permutation of p and q .

Sub-case iv: $d_p \neq \mathcal{U}, d_q \neq \mathcal{U}$ If $d_\alpha = \mathcal{U}$, then the regularity constraint is equivalent to

$$V''_{p,q}(l_p, d_p, l_q, d_q) \leq 2\lambda_{\text{smooth}}d_{\max}, \quad (15)$$

which is always satisfied. If $d_\alpha \neq \mathcal{U}$, then the regularity constraint is equivalent to

$$\min(|d_p - d_q|, d_{\max}) \leq \min(|d_p - d_\alpha|, d_{\max}) + \min(|d_\alpha - d_q|, d_{\max}), \quad (16)$$

which is always true because the truncated linear distance defines a metric on space of the depth labels.

Case 2: $l_p \neq l_q, l_p \neq l_\alpha, l_q \neq l_\alpha$ Each term $V''_{p,q}$ is constant, so the regularity constraint is trivially satisfied.

Case 3: $l_p = l_\alpha, l_p \neq l_q$ We have $V_{p,q}(l_p, d_p, l_q, d_q) = V_{p,q}(l_\alpha, d_\alpha, l_q, d_q)$ and the regularity constraint is satisfied because the remaining term $V_{p,q}(l_p, d_p, l_\alpha, d_\alpha)$ is by definition non-negative.

Case 4: $l_q = l_\alpha, l_p \neq l_q$ The proof is equivalent to the previous case after permutation of p and q .

Case 5: $l_p = l_q, l_p \neq l_\alpha$ The regularity constraint is equivalent to

$$V''_{p,q}(l_p, d_p, l_q, d_q) \leq 2\lambda_{\text{smooth}}d_{\max}, \quad (17)$$

which is always satisfied.

This completes the proof.

References

- [1] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *PAMI*, 26(9):1124–1137, 2004.
- [2] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 23(11):1222–1239, 2001.
- [3] J.-Y. Guillemaut, J. Kilner, and A. Hilton. Robust graph-cut scene segmentation and reconstruction for free-viewpoint video of complex dynamic scenes. In *ICCV*, 2009 (to appear).
- [4] V. Kolmogorov and R. Zabih. What energy function can be minimized via graph cuts? *PAMI*, 26(2):147–159, 2004.