Scene Text Recognition

Vincent BODIN & Thomas MOREAU

Scene text recognition refers to finding automatic ways of extracting text in pictures of every day life. As computer vision is more and more successful, scene text recognition has naturally become a significant issue nowadays. Unlike OCR which is more or less well understood and implemented, scene text recognition still needs many improvement to be considered as a powerful tool. We implemented paper [1] that explains how to extract words efficiently. A graphical model is introduced inside, with the creation of graph represented possible detected words, and algorithm TRW-S [2] is used to extract the optimal one.

1 Word recognition

This first part of the project has been done as a project for the course *Object recognition and computer vision.* It consists in:

1. Train classifiers: for every possible character in the English language, train a SVM with a RBF kernel (one-versus-all procedure). The database used is ICDAR 2003 [3]. Evaluate on a dictionary the frequency of occurrence of each pair of characters.
2. Recognizing character: performed via sliding window detection. For each possible window, compute the likelihood of each character and keep it if it has a confidence high enough.
3. Building words: build a graph of windows close enough in the x-direction. Attribute an energy to each possible word an energy depending on the lexicon prior.
4. Minimization of CRF energy: minimize this discrete energy using TRW-S algorithm [2].

PICTURES

2 TRW-S Algorithm

2.1 Preliminaries

Introduced by V. Kolmogorov [2], it is an algorithm that minimizes discrete energy on a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$:

$$E(\mathbf{x}|\theta) = \theta_c + \sum_{s\in\mathcal{V}} \theta_s(x_s) + \sum_{(s,t)\in\mathcal{E}} \theta_{st}(x_s, x_t) \tag{1}$$

In [4], an algorithm called *belief propagation* was described. It gives the exact optimum in the case of a tree, but fail to give good result otherwise because the algorithm might get stuck into a loop. It is a slight version of the sum-product algorithm in the sense that it could formally be obtained by switching the sum with max and the product with $+$ $((\mathbb{R}, \max, +)$ is a semi-ring). In [5], a new algorithm was introduced to deal with this NP-hard problem, called *max-product tree-reweighted message passing* (TRW). *Sequential tree-reweighted message passing* (TRW-S), [2] is better in the sense that it always find a 'local' maximum of the bound.

2.2 Algorithm

Given a graph $\mathcal{G}$ we extract a collection of trees $\mathcal{T}$ with a certain distribution $\rho^T, T \in \mathcal{T}$. We define θ^T as a vector θ compatible with T - *i.e.* a which is null outside T. Set:

$$\mathcal{A} = \{\theta \in \mathbb{R}^{d\times|\mathcal{T}|} | \theta^T \text{ for all } T \in \mathcal{T}\} \tag{2}$$

We denote by $\theta' \equiv \theta$ if they give the same energy E for all $\mathbf{x}$. Then the algorithm TRW-S consists in solving:

$$\max_{\theta\in\mathcal{A}, \sum_T \rho^T\theta^T \equiv \bar{\theta}} \Phi_\rho(\theta) = \max_{\theta\in\mathcal{A}, \sum_T \rho^T\theta^T \equiv \bar{\theta}} \sum_T \min_{x\in\mathcal{X}} \left\langle \theta^T, \phi(x) \right\rangle \tag{3}$$

Petit encart publicitaire

Pub pub

3 Partie 3

Blablabla partie3 colonne3 partie3

References

[1] C.V. Jawahar A. Mishra, K. Alahari. Top-down and bottom-up cues for scene text recognition. *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2687 – 2694, June 2012.

[2] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2006.

[3] Icdar 2003, robust word recognition dataset. `http://algoval.essex.ac.uk/icdar/Datasets.html`. Accessed: 2013-11-30.

[4] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann Publishers, 1988.

[5] T. S. Jaakkola M. J. Wainwright and A. S. Willsky. Map estimation via agreement on (hyper)trees: Message-passing and linear-programming approaches. *IEEE Transactions on Information Theory*, 51:3697 – 3717, November 2005.