

DISS. ETH NO. 21221

Transfer Schemes for Image Segmentation

A dissertation submitted to
ETH ZURICH

for the degree of
Doctor of Sciences

presented by
Daniel Kuettel
M.Sc. ETH in Computer Science
born October 12th, 1981
citizen of Gersau, Switzerland

accepted on the recommendation of
Prof. Dr. Vittorio Ferrari, examiner
Prof. Dr. Bastian Leibe, co-examiner
Prof. Dr. Marc Pollefeys, co-referent

2013

Abstract

The goal of figure-ground segmentation is to separate all foreground objects from the background. Many computer vision applications rely on figure-ground segmentation. Despite well-researched, it is still an unsolved problem.

In chapter 2, we present a technique for figure-ground segmentation in a fully supervised setting. The key idea is to transfer segmentation masks from training windows that are visually similar to windows in the test image. While most earlier works on annotation transfer are based on global descriptors, using windows as support regions facilitates transfer as they exhibit less variability than the whole image, and they enable to compose novel scenes using local parts from different training images. In the second stage, an energy-based segmentation model combines the transferred masks into a single coherent segmentation. This results in a fully automatic segmentation scheme, as opposed to existing interactive techniques based on similar energy functions. We experimentally demonstrate that our approach delivers state-of-the-art performance on several datasets (PASCAL VOC, Graz-02, Weizmann Horses).

In chapter 3, we extend our approach to segment the large-scale dataset ImageNet. It contains no figure-ground ground-truth, only a small fraction of images have bounding boxes, but all images are annotated with class labels organised in a semantic hierarchy. The key idea is to recursively exploit images segmented so far to guide the segmentation of new images. Our approach effectively takes advantage of the existing bounding-box annotations and of the semantic relations between classes. We show that it is computationally efficiently and accurately annotates 500k images over 577 classes.

In chapter 4, we present a technique for approximate minimization of two-label energy functions with higher-order potentials. We apply our

method to figure-ground segmentation and demonstrate experimentally that it enables to incorporate global shape priors.

In chapter 5, we propose a joint model that combines both global and local similarity for the task of assigning multiple semantic labels to a test image. We demonstrate that both similarities offer complementary information.

Zusammenfassung

Wir zeigen in Kapitel 2 eine Technik fuer Vordergrundsegmentierung mit kompletten Trainingsdaten. Die Grundidee is es, Segmentationsmasken von Trainingsfenstern zu aehnlichen Testfenstern zu transferieren. Normalerweise wurde das mit globaler Aehnlichkeit gemacht, aber Transfer mit Fenstern funktioniert besser, da ein Fenster weniger Variabilitaet enthaelt als ein ganzes Bild. Deshalb ist es damit moeglich, neue Szenen zu erklaren, indem man lokale Teile von verschiedenen Trainingsbildern benutzt. In einem zweiten Schritt kombiniert ein energiebasiertes Segmentierungsmodell die transferierten Masken zu einer zusammenhaengenden Segmentierung. Wir erreichen damit eine vollautomatische Segmentierung. Viele existierende Techniken funktionieren nur interaktiv. Wir demonstrieren experimentell das unsere Technik sehr gute Ergebnisse liefert auf mehreren Datensatzen (PASCAL VOC, Graz-02, Weizmann Horses).

In Kapitel 3 erweitern wir unsere Technik, um damit den riesigen Datensatz ImageNet zu segmentieren. Es enthaelt keine Vordergrundtrainingsdaten. Nur ein kleiner Teil der Bilder haben umschreibenden Rechtecke. Es haben jedoch alle Bilder einen Klassennamen und sind in einer semantischen Hierarchie organisiert. Die Grundidee ist es, rekursiv Bilder zu benutzen, die schon segmentiert wurden. Unsere Technik erlaubt es, auch umschreibende Rechtecke und semantische Beziehungen einfließen zu lassen. Wir zeigen, dass die Technik effizient ist, indem wir 500'000 Bilder mit 577 Klassen exakt segmentieren.

Im Kapitel 4 praesentieren wir eine Technik, um zwei-Label Energiefunktionen mit hoeheren Potentialen zu approximieren. Wir wenden unsere Methode fuer Vordergrundsegmentierung an und demonstrieren experimentell, dass sie erfolgreich globale Formprior modellieren kann.

In Kapitel 5 schlagen wir ein neues Modell vor, das gleichzeitig globale und lokale Ähnlichkeiten kombiniert, um semantische Label einem Testbild zuzuweisen. Wir demonstrieren, dass beide Ähnlichkeiten ergänzende Informationen bieten.