



Doctoral Thesis

## Large-Scale Image Recognition with Random Forests

**Author(s):**

Ristin-Kaufmann, Marko

**Publication Date:**

2015

**Permanent Link:**

<https://doi.org/10.3929/ethz-a-010536290> →

**Rights / License:**

[In Copyright - Non-Commercial Use Permitted](#) →

This page was generated automatically upon download from the [ETH Zurich Research Collection](#). For more information please consult the [Terms of use](#).

DISS. ETH NO. 22643

# **Large-Scale Image Recognition with Random Forests**

A thesis submitted to attain the degree of  
Doctor of Sciences of ETH Zurich  
(Dr. sc. ETH Zurich)

presented by

**Marko Ristin-Kaufmann**

MSc., ETH Zurich

born on November 22, 1984

citizen of Baar, Switzerland

accepted on the recommendation of

Prof. Dr. Luc Van Gool, examiner

Prof. Dr. Barbara Caputo, co-examiner

Prof. Dr. Jürgen Gall, co-examiner

2015

# Abstract

The recent advances in computer technology caused a revolution in digital imaging. The emergence of cheap good-quality cameras and inexpensive memory chips led to a virtual zero cost of image acquisition and storage. This, in its own part, ignited an expansion of visual data in many spheres, private and business alike, as billions of images are shared on a daily basis via social media sites like Facebook or Flickr. Manual organization of such a great pool of images becomes infeasible due to its size. As a remedy, we need to employ automatic methods of computer vision.

Earlier algorithms proposed by the computer vision community worked well on smaller datasets. However, new approaches need to be explored, since modern datasets present an increase of scale in multiple regards: they come with a larger quantity of images, as well as more complex ontologies, and require faster algorithms, if these datasets are to be processed in a reasonable time. In this thesis, we introduce three methods that address different aspects of image processing at a scale where thousands of object categories need to be recognized in millions of images. All three methods are developed within the framework of Random Forests, which is particularly suitable for such a large-scale setting.

As image collections grow in size, so does their variety increase. This diversification of visual content leads to more complex ontologies posing new challenges for recognition systems. Namely, as samples of new image classes appear, our classification models need to be extended to incorporate this novel knowledge. To avoid computational costs associated with re-training, these new classes have to be incrementally integrated into the models. To that end, we elaborate different update strategies for Random Forests. We experimentally show that forests updated in multiple batches achieve the same accuracy as a forest that was trained with the total training data.

To organize the diversity of large image collections, we not only have to introduce new categories, but we must also split existing ones into finer subcategories. However, this involves manual refinement of annotation for a large set of training samples, which is tedious and costly. In order to reduce the need for finer, but expensive annotations, we consider how coarse-labeled training samples can be leveraged to improve classification of fine subcategories by Random Forests. In a series of experiments we show that the accuracy of our forests almost matches the accuracy of the forests trained with fully annotated data.

Finally, when processing a large number of images, efficiency is of utter importance. All the more so in a detection pipeline which has to classify multiple subregions of an image to detect objects. In this thesis, we show how Random Forests can learn to predict subregions where objects can be expected based only on the local context of an image patch. Thus, processing time is reduced by examining fewer subregions, while we experimentally evaluate that our forests rarely miss relevant parts of the image.

# Zusammenfassung

Die rasante Entwicklung der Computertechnologie verursachte eine Revolution in der Welt der Digitalfotographie. Billige, jedoch qualitativ hochwertige Fotocameras und preiswerte Speicher ermöglichen Bildaufnahmen, die praktisch nichts mehr kosten. Das hat zu einer Ausweitung der bildlichen Daten in vielen Sphären des Alltagsleben geführt, sowohl in privaten wie auch in geschäftlichen, während täglich Milliarden von Fotos über Plattformen wie Facebook oder Flickr ausgetauscht werden. Diese Unmenge an Bildern kann man wegen ihrer grossen Zahl nur sehr schwer, wenn überhaupt, von Hand einordnen. Dafür müssen wir automatische Methoden der Bilderkennung einsetzen.

Frühere Algorithmen, die die Bilderkennungsgemeinschaft vorgeschlagen hat, funktionierten gut auf kleineren Datensätzen. Inzwischen benötigt man neue Ansätze, da die Datensätze indes gewachsen sind, und zwar in verschiedenen Dimensionen: Die Anzahl Bilder nahm zu, ihre Ontologien wurden komplizierter, und wir brauchen schnellere Algorithmen, um solche Datensätze schnell genug bearbeiten zu können. In dieser Dissertation werden wir drei Methoden einführen, die diese Problempunkte anschneiden und tausende von Kategorien in Millionen von Bildern erkennen müssen. Alle drei Methoden werden im Rahmen der Random Forests entwickelt, da Random Forests besonders für solche Probleme mit grossen Datenmengen geeignet sind.

Je grösser die Bildersammlung, desto abwechslungsreicher die Bilder. Diese Vielfalt führt zu komplizierteren Ontologien, die handkerum eine neue Herausforderung an das Erkennungssystem darstellen. Wenn Beispiele von neuen Bildkategorien auftauchen, muss man die Kategorisierungsmodelle um dieses neuartige Wissen ergänzen. Die Modelle jedes Mal neu zu berechnen ist sehr rechenaufwendig. Deswegen müssen diese neuen Kategorien schrittweise in das bestehende System integriert werden. Dafür erarbeiten wir verschiedene Erweiterungsstrategien von Random Forests. Die experimentellen Resultate zeigen, dass die schrittweise erweiterten Forests die gleiche Genauigkeit erreichen wie der Forest, den man von Grund auf mit allen Daten trainiert hat.

Wenn man die grossen Bildersammlungen organisiert, muss man nicht nur neue Kategorien einführen können, sondern muss sie auch in weitere Unterkategorien verfeinern können. Dafür muss man aber von Hand die Etikettierung von vielen Trainingsbeispielen entsprechend präzisieren, was aufwendig und teuer sein kann. Deswegen versuchen wir die Trainingsbeispiele mit gröberen Etiketten zu benutzen, um die erforderliche Menge an verfeinernden Etikettierungen zu verringern, und gleichzeitig die Einordnungsgenauigkeit von Forests über feine Kategorien hochzuhalten. In einer Reihen von Experimenten zeigen wir, dass die Forests, die gleichzeitig mit gröberen und feineren Etiketten trainiert wurden, fast die gleiche Genauigkeit erreichen, wie die Forests, die den Zugriff auf alle Beispiele mit feiner Etikettierung hatten.

Schlussendlich ist es sehr wichtig, dass man eine grosse Menge von Bildern effizient bearbeiten kann. Vor allem wenn man verschiedene Objekte in verschiedenen Teilen eines Bildes erkennen will. In dieser Dissertation zeigen wir, wie wir Random Forests einsetzen können, wenn man relevante Bildteile auf Grund des lokalen Bildkontexts vorschlagen will. Somit können wir die Bearbeitungszeit deutlich reduzieren, da weniger Bildregionen untersucht werden müssen, und zeigen Anhand von Experimenten, dass unsere Forests es selten versäumen, die wichtigen Bildteile zu erkennen.