INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team Willow

Models of visual object recognition and scene understanding

Paris - Rocquencourt

Theme : Vision, Perception and Multimedia Understanding

Activity Report

2010

Table of contents

1. Team

Research Scientists

Jean Ponce [Team Leader, Professor in the Département d'Informatique of École Normale Supérieure (ENS), and adjunct professor in the Department of Computer Science at the University of Illinois at Urbana-Champaign (UIUC), HdR]
Andrew Zisserman [Team Co-leader, Professor in the Engineering Department of the University of Oxford, and part-time professor at ENS, HdR]
Sylvain Arlot [Chargé de Recherches CNRS]
Jean-Yves Audibert [Chercheur at the Centre d'Enseignement et de Recherche en Technologies de l'Information et Systèmes (CERTIS) of the École Nationale des Ponts et Chaussées (ENPC), HdR]
Francis Bach ["Détaché" at INRIA from the Corps des Mines, HdR]
Ivan Laptev [Chargé de Recherches INRIA]
Josef Sivic [Chargé de Recherches INRIA]
Guillaume Obozinski [Ingénieur Expert de Recherche]

PhD Students

Louise Benoît
Y-Lan Boureau
Florent Couzinié-Devy
Olivier Duchenne
Loic Février
Toby Hocking
Rodolphe Jenatton
Armand Joulin
Augustin Lefèvre
Julien Mairal
Marc Sturzel
Oliver Whyte
Matthieu Solnon
Edouard Grave
Warith Harchaoui
Muhammad Muneeb Ullah
Vincent Delaitre

Post-Doctoral Fellows

Neva Cherniavsky
Bryan Russell
Karteek Alahari
Mikel Rodriguez
Nicolas Le Roux
Jan van Gemert

Visiting Scientists

Mladen Kolar
Fredo Durand
Alexei Efros
Ram Nevatia

Administrative Assistant

Cécile Espiègle

2. Overall Objectives

2.1. Statement

Object recognition —or, in a broader sense, scene understanding— is the ultimate scientific challenge of computer vision: After 40 years of research, robustly identifying the familiar objects (chair, person, pet), scene categories (beach, forest, office), and activity patterns (conversation, dance, picnic) depicted in family pictures, news segments, or feature films is still far beyond the capabilities of today's vision systems. On the other hand, truly successful object recognition and scene understanding technology will have a broad impact in application domains as varied as defense, entertainment, health care, human-computer interaction, image retrieval and data mining, industrial and personal robotics, manufacturing, scientific image analysis, surveillance and security, and transportation.

Despite the limitations of today's scene understanding technology, tremendous progress has been accomplished in the past ten years, due in part to the formulation of object recognition as a statistical pattern matching problem. The emphasis is in general on the features defining the patterns and on the algorithms used to learn and recognize them, rather than on the representation of object, scene, and activity categories, or the integrated interpretation of the various scene elements. WILLOW complements this approach with an ambitious research program explicitly addressing the representational issues involved in object recognition and, more generally, scene understanding.

Concretely, our objective is to develop geometric, physical, and statistical models for all components of the image interpretation process, including illumination, materials, objects, scenes, and human activities. These models will be used to tackle fundamental scientific challenges such as three-dimensional (3D) object and scene modeling, analysis, and retrieval; human activity capture and classification; and category-level object and scene recognition. They will also support applications with high scientific, societal, and/or economic impact in domains such as quantitative image analysis in science and humanities; film post-production and special effects; and video annotation, interpretation, and retrieval. Machine learning is a key part of our effort, with a balance of practical work in support of computer vision application, methodological research aimed at developing effective algorithms and architectures, and foundational work in learning theory.

WILLOW was created in 2007: It was recognized as an INRIA team in January 2007, and as an official project-team in June 2007. WILLOW is a joint research team between INRIA Paris Rocquencourt, Ecole Normale Supérieure (ENS) and Centre National de la Recherche Scientifique (CNRS). This year we have hired one new researcher: Guillaume Obozinski ("ingénieur expert", INRIA) has joined WILLOW in September 2010. In addition, we have hired three post-docs: Karteek Alahari, Mikel Rodriguez, Nicolas Le Roux, and five new PhD students: Edouard Grave, Warith Harchaoui, Muhammad Muneeb Ullah, Vincent Delaitre, Matthieu Solnon. Alexei Efros (Professor, Carnegie Mellon University, USA), Frédo Durand (MIT), and Ram Nevatia (USC, USA) visited WILLOW in 2010.

2.2. Highlights

- I. Laptev and J. Sivic (together with C. Schmid (INRIA Grenoble)) co-organized one week summer school on visual recognition and machine learning. http://www.di.ens.fr/willow/events/cvml2010/ The school has attracted 137 participants from 26 countries.
- Jean Ponce was awarded an Advanced ERC Grant
- The group is splitting into two on January 1st 2011 to create a new INRIA project-team called SIERRA. The new group and its interactions with WILLOW is described in section 6.7
- Sylvain Arlot will be teaching one of the two Cours Peccot at Collège de France in Spring 2011

3. Scientific Foundations

3.1. 3D object and scene modeling, analysis, and retrieval

This part of our research focuses on geometric models of specific 3D objects at the local (differential) and global levels, physical and statistical models of materials and illumination patterns, and modeling and retrieval of objects and scenes in large image collections. Our past work in these areas includes research aimed at recognizing rigid 3D objects in cluttered photographs taken from arbitrary viewpoints (Rothganger *et al.*, 2006), segmenting video sequences into parts corresponding to rigid scene components before recognizing these in new video clips (Rothganger *et al.*, 2007), and retrieval of particular objects and buildings from images and videos (Sivic and Zisserman, 2003) and (Philbin *et al.*, 2007). Our current research focuses on acquisition of detailed object models from multiple images and video streams, theoretical analysis of camera models, and object/scene retrieval.

3.1.1. High-fidelity image-based object and scene modeling.

We have recently developed multi-view stereopsis algorithms that have proven remarkably effective at recovering intricate details and thin features of compact objects and capturing the overall structure of large-scale, cluttered scenes. Some of the corresponding software (PMVS, http://grail.cs.washington.edu/software/pmvs/) is available for free for academics, and licensing negotiations with several companies are under way. Our current work extends this approach in two directions: the first one is theoretical, with a general formalism for modeling central and non-central cameras using the formalism and terminology of classical projective geometry (Section 6.1.5), while the second one is more applied, using our multi-view-stereo approach to model archaeological sites (Section 6.1.1), and using image matching to detect abandoned objects near roads (Sections 6.1.2 and 6.1.3).

3.1.2. Retrieval and modeling of objects and scenes in large image collections.

The goal of this research is to develop techniques for visual search and recognition of objects and scenes in large image collections. In addition, the goal is to also investigate novel applications of large scale recognition in other domains, such as image processing (e.g. image enhancement and restoration), computer graphics (novel scene synthesis, visualization), 3D reconstruction, or visual localization.

We have introduced a geometric Latent Dirichlet Allocation (gLDA) model for unsupervised modeling of unstructured image collections (Section 6.2.1). Further, we achieved a significant reduction of quantization errors in large scale retrieval by learning a better local descriptor from large amounts of automatically generated matched/non-matched training data (Section 6.2.2).

In the context of large scale place recognition in structured, geo-referenced image databases, we have developed (Section 6.2.3) an approach for avoiding confusing features (such as trees or road markings), and introduced a localization approach based on matching linear combinations of views (Section 6.2.4) . Both methods lead to significant reductions of localization errors.

In terms of applications, we have investigated features and scene category recognition techniques for synthesizing novel scenes and navigating large collections of still images (Sections 6.2.6 and 6.2.5) . In the direction of image restoration and enhancement, we have developed a new geometrical model for non-uniform blurs caused by camera shake (Section 6.2.7) .

3.2. Category-level object and scene recognition

The objective in this core part of our research is to learn and recognize quickly and accurately thousands of visual categories, including materials, objects, scenes, and broad classes of temporal events, such as patterns of human activities in picnics, conversations, etc. The current paradigm in the vision community is to model/learn one object category (read 2D aspect) at a time. If we are to achieve our goal, we have to break away from this paradigm, and develop models that account for the tremendous variability in object and scene appearance due to texture, material, viewpoint, and illumination changes within each object category, as well as the complex and evolving relationships between scene elements during the course of normal human activities. Our current work focuses on the following problems:

3.2.1. Learning image and object models.

Learning sparse representations of images has been the topic of much recent research. It has been used for instance for image restoration (e.g., Mairal et al., 2007) and it has been generalized to discriminative image understanding tasks such as texture segmentation, category-level edge selection and image classification (Mairal et al., 2008). As discussed in Section 6.5.5, we have developed fast and scalable optimization methods for learning the sparse image representations, and developed a software called SPAMS (SPArse Modelling Software) presented in Section 5.2. The work of J. Mairal is summarized in his thesis (Section 6.5.7).

In addition, we have investigated methods to learn better mid-level features for recognition, including novel discriminative codebook learning algorithm using supervised backpropagation (Section 6.3.1) and theoretical analysis of local feature pooling (Section 6.3.2).

3.2.2. Category-level object/scene recognition and segmentation

Another significant strand of our research has focused on the extremely challenging goals of category-level object/scene recognition and segmentation. Towards these goals, we have developed a (i) discriminative clustering approach for image co-segmentation (Section 6.5.2) and (ii) context-dependent kernels for object classification (Section 6.5.4).

3.3. Human activity capture and classification

From a scientific point of view, visual action understanding is a computer vision problem that has received little attention so far outside of extremely specific contexts such as surveillance or sports. Current approaches to the visual interpretation of human activities are designed for a limited range of operating conditions, such as static cameras, fixed scenes, or restricted actions. The objective of this part of our project is to attack the much more challenging problem of understanding actions and interactions in unconstrained video depicting everyday human activities such as in sitcoms, feature films, or news segments. The recent emergence of automated annotation tools for this type of video data (Everingham, Sivic, Zisserman, 2006; Laptev, Marszałek, Schmid, Rozenfeld, 2008) means that massive amounts of labelled data for training and recognizing action models will at long last be available.

3.3.1. Recognition of characters and their traits in video

We pursue the goal of automatically characterizing people in realistic video data. Along this direction, we have earlier explored automatic naming of characters in video by learning character-specific classifiers from the video data and associated scripts. In the recent extension of this work we consider semi-supervised learning of facial attributes such as gender, age and race. While facial attributes have been previously addressed in still images, we demonstrate improvements of facial attribute classification in video when automatically annotating faces in video training data.

3.3.2. Weakly-supervised learning and annotation of human actions in video

We aim to leverage the huge amount of video data using readily-available annotations in the form of video scripts. Scripts, however, often provide only imprecise and incomplete information about the video. We address this problem with weakly-supervised learning techniques both at the text and image levels. To this end we recently explored automatic mining of scene and action categories. We are currently extending this work towards learning a large pool of actions with the goal of using them as prmitives when learning and recognizing new actions categories.

3.3.3. Descriptors for video representation

Video representation has a crucial role for recognizing human actions and other components of a visual scene. Our work in this domain aims to develop generic methods for representing video data based on realistic assumptions. We explore the ways of enriching standard bag-of-feature representations with the higher-level information on objects and scenes pre-learned on related tasks. We also aim to capture higher level structural relations between humans, objects and scenes. Along these strands we are particularly investigating long-term temporal relations in the video which, for example, enable reasoning about the depth ordering of objects as well as the temporal ordering actions in dynamical scenes.

3.3.4. Crowd characterization in video

Human crowds are characterized by distinct visual appearance and require appropriate tools for their analysis. In our work we develop generic methods for crowd analysis in video aiming to address multiple tasks such as (i) crowd density estimation and localization, (ii) characterization and recognition of crowd behaviours (e.g a person running against the crowd flow) as well as (iii) detection and tracking of individual people in the crowd. We address the challenge of analyzing crowds under the large variation in crowd density, video resolution and scene structure.

3.3.5. Action recognition in still images

Recognition of human actions is usually addressed in the scope of video interpretation. Meanwhile, common human actions such as "reading a book", "playing a guitar" or "writing notes" also provide a natural description for many still images. Motivated by the potential impact of recognizing actions in still images, we address recognition of human actions in consumer photographs. We have so far studied performance of several state-of-the-art visual recognition methods applied to existing datasets and our newly collected dataset with 968 Flickr images and seven classes of human actions.

3.4. Machine learning

A large portion of research in computer vision involves increasingly more refined machine learning techniques. Part of our effort focuses on a careful integration of state of the art machine learning in vision, but we also pursue several lines of research aiming at providing a better understanding of the fundamental ideas leading to efficient learning algorithms.

3.4.1. Machine learning for computer vision.

Significant success in vision has been obtained by the direct use of off-the-shelf machine learning techniques, such as kernel methods (support vector machines for example) and probabilistic graphical models. However, in order to achieve the level of performance that we aim for, a more careful integration of machine learning and computer vision algorithmic and theoretical frameworks is needed.

Sparse coding provides a power tool to tackle a broad array of problem in image processing including denoising, inpainting, texture synthesis, etc. We developed efficient algorithmic ideas based on online optimization to solve problems involving very large images, and very large databases of images (6.5.5) and were able to use the proposed methodology on a large scale for various tasks (6.5.6, 6.5.7) We also applied the proposed algorithmic principle to the related setting of topic models (6.6.4).

We pursued a line of work on learning problems for combinatorial settings such as problems with geometric constraints, including template alignments via matching (6.5.1, 6.5.4), segmentation with graph cuts(6.5.2), or latent classes (6.5.3).

3.4.2. Algorithms and Learning theory.

Common features encountered when using learning techniques in computer vision are (i) high dimensionality and (ii) complexity of the modelisation. Sparse methods allow to avoid the curse of dimensionality to some extent, lead to interpretable models and good performance. We are currently exploring structured sparse methods, where the idea is to introduce some prior knowledge into a sparse inference problem, for computational reasons or to improve interpretability and predictive performance (see sections 6.6.3, 6.6.7,6.6.8, 6.6.9).

We pursued several lines of work on cross validation and data-driven calibration procedures (6.6.2), on sequential prediction (6.6.10,6.6.11, 6.6.12, 6.6.13), and on learning rates for adaptive procedures (6.6.1) and smooth convex losses (6.6.5).

4. Application Domains

4.1. Introduction

We believe that foundational modeling work should be grounded in applications. This includes (but is not restricted to) the following high-impact domains.

4.2. Quantitative image analysis in science and humanities

We plan to apply our 3D object and scene modeling and analysis technology to image-based modeling of human skeletons and artifacts in anthropology, and large-scale site indexing, modeling, and retrieval in archaeology and cultural heritage preservation. Most existing work in this domain concentrates on image-based rendering—that is, the synthesis of good-looking pictures of artifacts and digs. We plan to focus instead on quantitative applications. A first effort in this area has been a collaboration with the Getty Conservation Institute in Los Angeles, aimed at the quantitative analysis of environmental effects on the hieroglyphic stairway at the Copan Maya site in Honduras. We are now pursuing a larger-scale project involving the archaeology laboratory at ENS and focusing on image-based artifact modeling and decorative pattern retrieval in Pompeii. This new effort is part of the MSR-INRIA project mentioned earlier and that will be discussed further later in this report.

4.3. Video Annotation, Interpretation, and Retrieval

Both specific and category-level object and scene recognition can be used to annotate, augment, index, and retrieve video segments in the audiovisual domain. The Video Google system developed by Sivic and Zisserman (2005) for retrieving shots containing specific objects is an early success in that area. A sample application, suggested by discussions with Institut National de l'Audiovisuel (INA) staff, is to match set photographs with actual shots in film and video archives, despite the fact that detailed timetables and/or annotations are typically not available for either medium. Automatically annotating the shots is of course also relevant for archives that may record hundreds of thousands of hours of video. Some of these applications will be pursued in our MSR-INRIA project, in which INA is one of our partners.

5. Software

5.1. Change-Point Detection via Cross-Validation

A family of change-point detection procedures via cross-validation was proposed in the paper [7], where the goal is to detect changes in the mean of a signal that can be heteroscedastic. This software is a Matlab package allowing to perform the new procedures and to generate synthetic data as in the paper. This package is provided free for non-commercial use under the terms of the GNU General Public License. It is publicly available at the url http://www.di.ens.fr/~arlot/code/CHPTCV.htm.

5.2. SPArse Modeling Software (SPAMS)

SPAMS v2.0 was released in November 2010 (v1.0 was released in September 2009). It is an optimization toolbox composed of a set of binaries implementing algorithms to address various machine learning and signal processing problems involving

- Dictionary learning and matrix factorization (NMF, sparse PCA, ...)
- Solving sparse decomposition problems with LARS, coordinate descent, OMP, SOMP, proximal methods
- Solving structured sparse decomposition problems (ℓ_1/ℓ_2, ℓ_1/ℓ_∞, sparse group lasso, tree-structured regularization, structured sparsity with overlapping groups,...).

The software and its documentation are available at http://www.di.ens.fr/willow/SPAMS/.

5.3. Non-uniform Deblurring for Shaken Images

A package of Matlab code for removing non-uniform camera shake blur from a single blurry image. The algorithm is described in [47]. The package is publicly available at http://www.di.ens.fr/willow/research/deblurring/.

5.4. Local dense and sparse space-time features

A package with Linux binaries implementing extraction of local space-time features in video. The code supports feature extraction at Harris3D points, on a dense space-time grid as well as at user-supplied space-time locations. The package is publicly available at http://www.di.ens.fr/~laptev/download/stip-2.0-linux.zip.

5.5. Detecting a road

A package of Matlab code for detecting in a single image an arbitrary road, that may not be well-paved, or have clearly delineated edges, or some a priori known color or texture distribution. Related publications and code can be found in the project web page http://bmi.osu.edu/~hkong/Road_Detection.html.

6. New Results

6.1. High-fidelity image- and video-based modeling

6.1.1. Quantitative image analysis for archeology (B. Russell, J. Ponce, J. Sivic, joint work with H. Dessales, ENS Archeology laboratory)

Accurate indexing and alignment of images is an important problem in computer vision. A successful system would allow a user to retrieve images with similar content to a query image, along with any information associated with the image. Prior work has mostly focused on techniques to index and match photographs depicting particular instances of objects or scenes (e.g. famous landmarks, commercial product labels, etc.). This has allowed progress on tasks, such as the recovery of a 3D reconstruction of the depicted scene.

However, there are many types of images that cannot be accurately aligned. For instance, for many locations there are drawings and paintings made by artists that depict the scene. Matching and aligning photographs, paintings, and drawings is extremely difficult due to various distortions that can arise. Examples include perspective and caricature distortions, along with errors that arise due to the difficulty of drawing a scene by hand.

In this project, we seek to index and align a database of images, paintings, and drawings. The focus of our work is the Championnet house in the Roman ruins at Pompeii, Italy. Given an alignment of the images, paintings, and drawings, we wish to explore tasks that are of interest to archaeologists and curators who wish to study and preserve the site. Example applications include: (i) digitally restoring paintings on walls where the paintings have disappeared over time due to erosion, (ii) geometrically reasoning about the site over time through the drawings, (iii) indexing and searching patterns that exist throughout the site.

To date, we have visited the site in Pompeii and photographed the rooms of interest. An initial dense 3D reconstruction has been achieved from 585 photographs using existing photometric multi-view stereo methods. Figure 1 shows some of the captured photographs and snapshots of the 3D reconstruction of the site. Notice that the 3D reconstruction captures much detail of the walls and structures.

Next, we have obtained initial results on coarse alignment of paintings with the 3D model of the site. This is achieved by matching to virtual viewpoints that are uniformly sampled across the 3D model. The result is a viewpoint that is sufficiently close to the painting viewpoint, where the depicted scene objects in the painting are close to their 3D model projection.

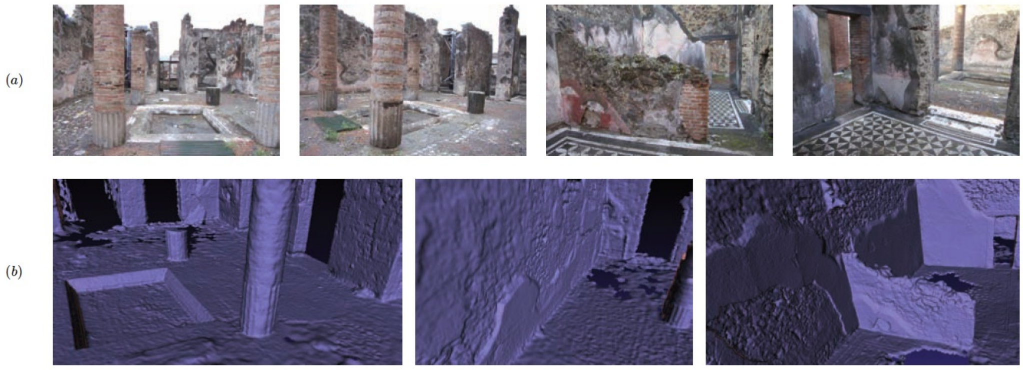

Figure 1. (a) Example photographs captured of the Pompeii site (563 photographs are used in total). (b) Rendered viewpoints of the recovered 3D model. Notice the fine-level details that are captured by the model.

Currently we are exploring different techniques to refine the obtained viewpoints and align the paintings, and drawings with the 3D model. We hope to submit results from our research to a conference in Spring 2011.

6.1.2. Detecting abandoned objects with a moving camera (H. Kong, J.-Y. Audibert, J. Ponce)

We design a novel framework for detecting non-flat abandoned objects by matching a reference and a target video sequences. The reference video is taken by a moving camera when there is no suspicious object in the scene. The target video is taken by a camera following the same route and may contain extra objects. The objective is to find these objects. GPS information is used to roughly align the two videos and find the corresponding frame pairs. Based on the GPS alignment, four simple but effective ideas are proposed to achieve the objective: an inter-sequence geometric alignment based on homographies, which is computed by a modified RANSAC, to find all possible suspicious areas, an intra-sequence geometric alignment to remove false alarms caused by high objects, a local appearance comparison between two aligned intra-sequence frames to remove false alarms in flat areas, and a temporal filtering step to confirm the existence of suspicious objects. Experiments on fifteen pairs of videos show the promise of the proposed method.

This work is a follow-up to one of our 2009 CVPR conference paper, and resulted in a journal publication [17].

6.1.3. General road detection from a single image (H. Kong, J.-Y. Audibert, J. Ponce)

Given a single image of an arbitrary road, that may not be well-paved, or have clearly delineated edges, or some a priori known color or texture distribution, is it possible for a computer to find this road? This paper addresses this question by decomposing the road detection process into two steps: the estimation of the vanishing point associated with the main (straight) part of the road, followed by the segmentation of the corresponding road area based on the detected vanishing point. The main technical contributions of the proposed approach are a novel adaptive soft voting scheme based on a local voting region using high-confidence voters, whose texture orientations are computed using Gabor filters, and a new vanishing-point-constrained edge detection technique for detecting road boundaries. The proposed method has been implemented, and experiments with 1003 general road images demonstrate that it is effective at detecting road regions in challenging conditions.

This work is a follow-up to one of our 2009 CVPR conference paper, and resulted in a journal publication [16].

6.1.4. Accurate, Dense, and Robust Multi-View Stereopsis (J. Ponce, joint work with Y. Furukawa)

We have proposed a novel algorithm for multiview stereopsis that outputs a dense set of small rectangular patches covering the surfaces visible in the images. Stereopsis is implemented as a match, expand, and filter procedure, starting from a sparse set of matched keypoints, and repeatedly expanding these before using visibility constraints to filter away false matches. The keys to the performance of the proposed algorithm are effective techniques for enforcing local photometric consistency and global visibility constraints. Simple but effective methods are also proposed to turn the resulting patch model into a mesh which can be further refined by an algorithm that enforces both photometric consistency and regularization constraints. The proposed approach automatically detects and discards outliers and obstacles, and does not require any initialization in the form of a visual hull, a bounding box, or valid depth ranges. We have tested our algorithm on various datasets including objects with fine surface details, deep concavities, and thin structures, outdoor scenes observed from a restricted set of viewpoints, and crowded scenes where moving obstacles appear in front of a static structure of interest. A quantitative evaluation on the Middlebury benchmark (Seitz et al., 2006) shows that the proposed method outperforms all others submitted so far for four out of the six datasets.

This work has resulted in a PAMI publication [11]. A US patent application for the corresponding software is pending.

6.1.5. Admissible Map Models of Linear Cameras (J. Ponce, joint work with G. Batog and X. Goaoc, INRIA Nancy Grand Est, and M. Lavandier, Université de Poitiers)

We are continuing our investigation of general linear camera models. In particular, we have introduced a complete analytical characterization of a large class of central and non-central imaging devices dubbed *linear cameras* by Ponce (2009). Pajdla (2002) has shown that a subset of these, the oblique cameras, can be modelled by a certain type of linear map. We have obbtained a full tabulation of all *admissible* maps that induce cameras in the general sense of Grossberg and Nayar (2005), and showm that these cameras are exactly the linear ones. Combining these two models with a new notion of intrinsic parameters and normalized coordinates for linear cameras has allowed us to give simple analytical formulas for direct and inverse projection. We have also shown that the epipolar geometry of any two linear cameras can be characterized by a *fundamental matrix* whose size is at most 6×6 when the cameras are uncalibrated, or by an *essential matrix* of size at most 4×4 when their internal parameters are known. Similar results hold for trinocular constraints. A physical prototype of a *parabolic* camera has also been constructed.

This work has resulted in a CVPR'10 publication [25].

6.2. Retrieval and modeling of objects and scenes in large image collections

6.2.1. Geometric Latent Dirichlet Allocation on a Matching Graph for Large-Scale Image Datasets (J. Sivic and A. Zisserman, joint work with J. Philbin, Oxford University)

Given a large-scale collection of images we would like to be able to conceptually group together images taken of the same place, of the same thing, or of the same person.

To achieve this, we introduce the Geometric Latent Dirichlet Allocation (gLDA) model for unsupervised particular object discovery in unordered image collections. This explicitly represents documents as mixtures of particular objects or facades, and builds rich latent topic models which incorporate the identity and locations of visual words specific to the topic in a geometrically consistent way. Applying standard inference techniques to this model enables images likely to contain the same object to be probabilistically grouped and ranked.

Additionally, to reduce the computational cost of applying our model to large datasets, we describe a scalable method that first computes a matching graph over all the images in a dataset. This matching graph connects images that contain the same object and rough image groups can be mined from this graph using standard clustering techniques. The gLDA model can then be applied to generate a more nuanced representation of the data. We also discuss how "hub images" (images representative of an object or landmark) can easily be extracted from our matching graph representation.

We evaluate our techniques on the publicly available Oxford buildings dataset (5K images) and show examples of objects automatically mined from this dataset. The methods are evaluated quantitatively on this dataset using a ground truth labeling for a number of Oxford landmarks. To demonstrate the scalability of the matching graph method, we show qualitative results on two larger datasets of images taken of the Statue of Liberty (37K images) and Rome (1M+ images).

The project resulted in a publication [20].

6.2.2. Descriptor learning for efficient retrieval (J. Sivic and A. Zisserman, joint work with J. Philbin, Google and M. Isard, Microsoft)

Many visual search and matching systems represent images using sparse sets of “visual words”: descriptors that have been quantized by assignment to the best-matching symbol in a discrete vocabulary. Errors in this quantization procedure propagate throughout the rest of the system, either harming performance or requiring correction using additional storage or processing. This paper directly addresses these quantization errors by learning a projection from descriptor space to a new Euclidean space in which standard clustering techniques are more likely to assign matching descriptors to the same cluster, and non-matching descriptors to different clusters.

To learn the projection function we develop a novel cost function incorporating margin based losses, which separates matching descriptors from two classes of non-matching descriptors; and show that a non-linear projection function gives better performance than the linear methods previously used in computer vision systems. We also develop a simple automatic procedure to generate large amounts of matching/non-matching descriptor training data from a corpus of unlabeled images; and show that stochastic gradient methods can be successfully used for optimizing the cost function over such massive amounts of training data.

For the case of particular object retrieval, we demonstrate impressive gains in performance on a ground truth dataset: our learnt 32-D descriptor without spatial re-ranking outperforms a baseline method using 128-D SIFT descriptors with spatial re-ranking.

The project resulted in a publication [43].

6.2.3. Avoiding confusing features in place recognition (J. Sivic, in collaboration with J. Knopp, CTU Prague / KU Leuven, and T. Pajdla, CTU Prague)

We seek to recognize the place depicted in a query image using a database of “street side" images annotated with geolocation information. This is a challenging task due to changes in scale, viewpoint and lighting between the query and the images in the database. One of the key problems in place recognition is the presence of objects such as trees or road markings, which frequently occur in the database and hence cause significant confusion between different places. As the main contribution, we show how to avoid features leading to confusion of particular places by using geotags attached to database images as a form of supervision. We develop a method for automatic detection of image-specific and spatially-localized groups of confusing features, and demonstrate that suppressing them significantly improves place recognition performance while reducing the database size. We show the method combines well with the state of the art bag-of-features model including query expansion, and demonstrate place recognition that generalizes over wide range of viewpoints and lighting conditions. Results are shown on a geotagged database of over 17K images of Paris downloaded from Google Street View. Example results are shown in figure 2.

The project resulted in a publication [39].

6.2.4. Read between the views: image-based localization by matching linear combinations of views (J. Sivic, joint work with A. Torii, Tokyo Institute of Technology and T. Pajdla, CTU in Prague)

We seek to predict the GPS location of a query image given a database of images with known GPS locations. We formulate this task as a regression problem. Both the query image and images in the database are described using the efficient bag-of-features representation. The goal is to obtain a mapping from the bag-of-feature

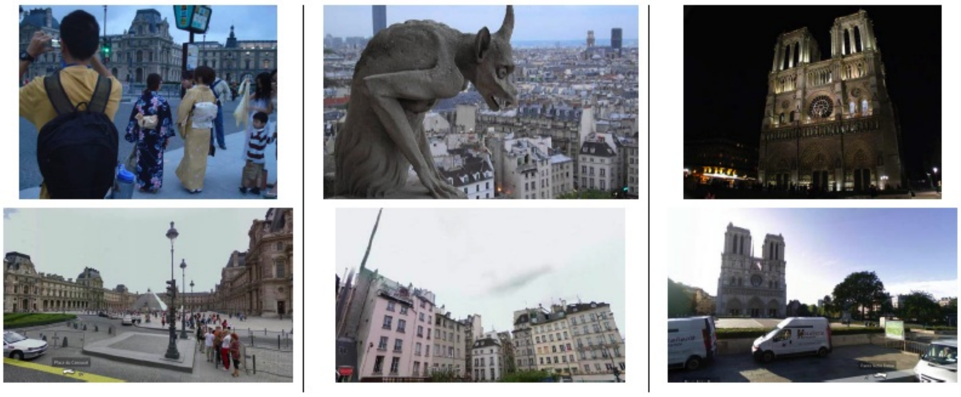

Figure 2. Examples of visual place recognition results. Given a query image (top) of an unknown place, the goal is to find an image from a geotagged database of street side imagery (bottom), depicting the same place as the query.

representation of the query image to its predicted position on the map. The main contribution of this paper is a two stage regression algorithm that takes into account the spatial organization of images in the database. In the first stage, we find the best matching pair of database images considering *linear combinations* of bag-of-feature vectors of spatially close-by images on the map. In the second stage, we predict the location of the query from the GPS locations of images in the matched pair. By considering interpolated views *in the feature space* during matching, this approach enables generalization to unseen views while reducing the database size. We demonstrate the proposed method outperforms other commonly used matching approaches. Results are shown on a database of 8,999 omni-directional Google Street-view images of Pittsburgh.

This work is under submission [45].

6.2.5. Infinite Images: Creating and Exploring a Large Photorealistic Virtual Space (J. Sivic, joint work with B. Kaneva, MIT, A. Torralba, MIT, S. Avidan, Adobe Research, W.T. Freeman, MIT)

We present a system for generating “infinite” images from large collections of photos by means of transformed image retrieval. Given a query image, we first transform it to simulate how it would look if the camera moved sideways and then perform image retrieval based on the transformed image. We then blend the query and retrieved images to create a larger panorama. Repeating this process will produce an “infinite" image. The transformed image retrieval model is not limited to simple 2D left/right image translation, however, and we show how to approximate other camera motions like rotation and forward motion/zoom-in using simple 2D image transforms. We represent images in the database as a graph where each node is an image and different types of edges correspond to different types of geometric transformations simulating different camera motions. Generating infinite images is thus reduced to following paths in the image graph. Given this data structure we can also generate a panorama that connects two query images, simply by finding the shortest path between the two in the image graph. We call this option the “image taxi". Our approach does not assume photographs are of a single real 3D location, nor that they were taken at the same time. Instead, we organize the photos in themes, such as city streets or skylines and synthesize new virtual scenes by combining images from distinct but visually similar locations. There are a number of potential applications to this technology. It can be used to generate long panoramas as well as content aware transitions between reference images or video shots. Finally, the image graph allows users to interactively explore large photo collections for ideation, games, social interaction and artistic purposes.

The project resulted in a publication [15].

6.2.6. *Matching and predicting street level images (J. Sivic, in collaboration with B. Kaneva, MIT, S. Avidan, Adobe, W. T. Freeman, MIT, and A. Torralba, MIT)*

The paradigm of matching images to a very large dataset has been used for numerous vision tasks and appears to be a powerful paradigm. If the image dataset is large enough, one can expect to find good matches of almost any image to the database, allowing label transfer (Liu et al. 2009, Berg and Malik 2005), and image editing or enhancement (Hays and Efros 2007, Dale et al. 2009). Users of this approach will want to how many images are required, and what features to use for finding semantic relevant matches. Furthermore, for navigation tasks or to exploit context, users will want to know the predictive quality of the dataset: can we predict the image that would be seen under changes in camera position?

We address these questions in detail for one category of images: street level views. We have a dataset of images taken from an enumeration of positions and viewpoints within Pittsburgh. We evaluate how well we can match those images, using images from non-Pittsburgh cities, and how well we can predict the images that would be seen under changes in camera position. We compare performance for these tasks for five different feature sets, finding a feature set that outperforms the others (HOG). We used Amazon Mechanical Turk workers to rank the matches and predictions of different algorithm conditions by comparing each one to the selection of a random image. This approach can evaluate the efficacy and optimal parameter settings for this general approach for other image categories and tasks.

The project resulted in a publication [37].

6.2.7. *Non-uniform Deblurring for Shaken Images (O. Whyte, J. Sivic, A. Zisserman and J. Ponce)*

We argue that blur resulting from camera shake is mostly due to the 3D rotation of the camera, causing a blur that can be significantly non-uniform across the image. How- ever, most current deblurring methods model the observed image as a convolution of a sharp image with a uniform blur kernel. We propose a new parametrized geometric model of the blurring process in terms of the rotational velocity of the camera during exposure. We apply this model in the context of two different algorithms for camera shake removal: the first uses a single blurry image (blind deblurring), while the second uses both a blurry image and a sharp but noisy im- age of the same scene. We show that our approach makes it possible to model and remove a wider class of blurs than previous approaches, and demonstrate its effectiveness with experiments on real images. Example result is shown in figure 3.

The project resulted in a publication [47].

6.3. Learning image and object models

6.3.1. *Learning Mid-Level Features For Recognition (Y-Lan Boureau, F.Bach and J. Ponce, together with Y. LeCun (NYU))*

In [27], we aim to facilitate the design of better recognition architectures with the following contributions:

- Experimental evaluation of many different intermediate coding schemes, which leads to state-of-the-art results on two recognition benchmarks
- Novel discriminative codebook learning algorithm using supervised backpropagation
- Theoretical and experimental investigation into the much better linear discrimination performance of max pooling compared to average pooling

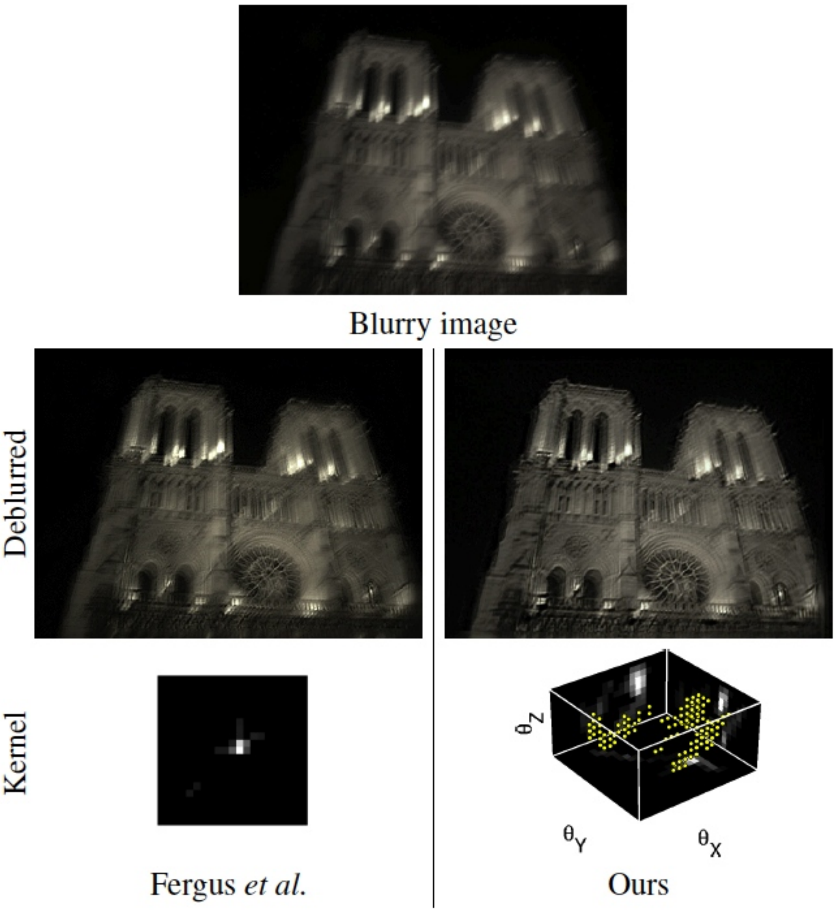

Figure 3. ***Blind deblurring of real camera shake, example 1.*** *The result of blind deblurring on a real camera shake image, captured with a shutter speed of $\frac{1}{2}$ second, using the algorithm of Fergus et al. and our non-uniform approach. Our approach is able to recover a useful kernel and a good deblurred image, while the uniform algorithm of Fergus et al. fails to find a meaningful kernel. The rotational kernel visualized in the right-hand column shows the non-zero kernel elements plotted as points in the 3D rotational parameter space ($\theta_X, \theta_Y, \theta_Z$). Each of the cuboid's faces shows the projection of the kernel onto that face. Note that our estimated rotational kernel has a significant in-plane component (non-zeros over many values of θ_Z).*

6.3.2. A Theoretical Analysis of Feature Pooling in Visual Recognition (Y-Lan Boureau and J. Ponce, together with Y. LeCun (NYU))

The work in [27] is further extended in [28], where we achieve a better understanding of pooling by:

- Extensively analysing the discriminative powers of different pooling operations
- Discriminating several factors affecting pooling performance, including smoothing and sparsity of the features
- Unifying several popular pooling schemes as part of a single continuum

6.4. Human activity capture and classification

6.4.1. Recognizing human actions in still images: a study of bag-of-features and part-based representations (V. Delaitre, I. Laptev and J. Sivic)

Recognition of human actions is usually addressed in the scope of video interpretation. Meanwhile, common human actions such as “reading a book”, “playing a guitar” or “writing notes” also provide a natural description for many still images. In addition, some actions in video such as “taking a photograph” are static by their nature and may require recognition methods based on static cues only. Motivated by the potential impact of recognizing actions in still images and the little attention this problem has received in computer vision so far, we address recognition of human actions in consumer photographs. We construct a new dataset with seven classes of actions in 968 Flickr images representing natural variations of human actions in terms of camera view-point, human pose, clothing, occlusions and scene background (examples shown in figure 4). We study action recognition in still images using the state-of-the-art bag-of-features methods as well as their combination with the part-based Latent SVM approach of Felzenszwalb et al. (PAMI 2009). In particular, we investigate the role of background scene context and demonstrate that improved action recognition performance can be achieved by (i) combining the statistical and part-based representations, and (ii) integrating person-centric description with the background scene context. We show results on our newly collected dataset of seven common actions as well as demonstrate improved performance over existing methods on the datasets of Gupta et al. (PAMI 2009) and Yao and Fei-Fei (CVPR 2010).

This work resulted in a publication [30].

6.4.2. Improving Bag-of-Features Action Recognition with Non-local Cues (M.M. Ullah, S.N. Parizi and I. Laptev)

Local space-time features have recently shown promising results within Bag-of-Features (BoF) approach to action recognition in video. Pure local features and descriptors, however, provide only limited discriminative power implying ambiguity among features and sub-optimal classification performance. In this work, we propose to disambiguate local space-time features and to improve action recognition by integrating additional non-local cues with BoF representation. For this purpose, we decompose video into region classes and augment local features with corresponding region-class labels. In particular, we investigate unsupervised and supervised video segmentation using (i) motion-based foreground segmentation, (ii) person detection, (iii) static action detection and (iv) object detection. While such segmentation methods might be imperfect, they provide complementary region-level information to local features. We demonstrate how this information can be integrated with BoF representations in a kernel-combination framework. We evaluate our method on the recent and challenging Hollywood-2 action dataset and demonstrate significant improvements.

This project resulted in a publication [46].

6.4.3. Semi-supervised learning of facial attributes in video (N. Cherniavsky, I. Laptev, J. Sivic and Andrew Zisserman)

In this work we investigate a weakly-supervised approach to learning facial attributes of humans in video. Given a small set of images labeled with attributes and a much larger unlabeled set of video tracks (see

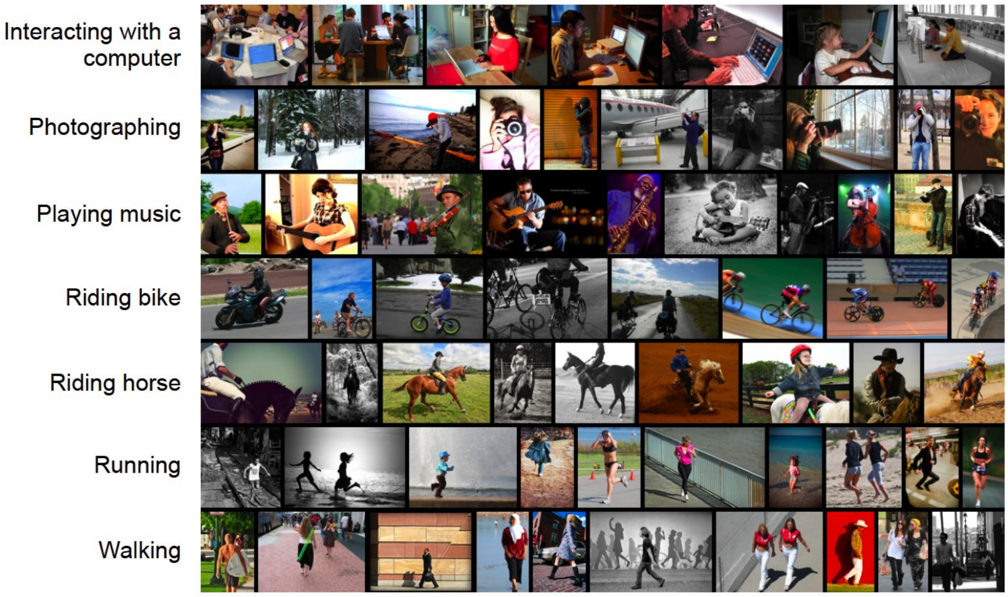

Figure 4. Example images from the newly constructed dataset with seven human action classes collected from Flickr. Note the natural and challenging variations in the camera view-point, clothing of people, occlusions, object appearance and scene layout present in the consumer photographs.

figure 5), we train a classifier to recognize these attributes in video data. We make two contributions. First, we show that training on video data improves classification performance over training on images alone. Second, and more significantly, we show that tracks in video provide a natural mechanism for generalizing training data – in this case to new poses, lighting conditions and expressions. The advantage of our method is demonstrated on the classification of gender and age attributes in the movie "Love, Actually". We show that the semi-supervised approach adds a significant performance boost, for example for gender increasing average precision from 0.75 on static images alone to 0.85.

This project resulted in a publication [29].

6.4.4. View-Independent Action Recognition from Temporal Self-Similarities (I. Laptev in collboration with I.N. Junejo, E. Dexter, and P. Pérez)

In this work we address recognition of human actions under view changes. We explore self-similarities of action sequences over time and observe the striking stability of such measures across views. Building upon this key observation, we develop an action descriptor that captures the structure of temporal similarities and dissimilarities within an action sequence. Despite this temporal self-similarity descriptor not being strictly view-invariant, we provide intuition and experimental validation demonstrating its high stability under view changes. Self-similarity descriptors are also shown stable under performance variations within a class of actions, when individual speed fluctuations are ignored. If required, such fluctuations between two different instances of the same action class can be explicitly recovered with dynamic time warping, as will be demonstrated, to achieve cross-view action synchronization. More central to present work, temporal ordering of local self-similarity descriptors can simply be ignored within a bag-of-features type of approach. Sufficient action discrimination is still retained this way to build a view-independent action recognition system. Interestingly, self-similarities computed from different image features possess similar properties and can be used in a complementary fashion. Our method is simple and requires neither structure recovery nor multi-view

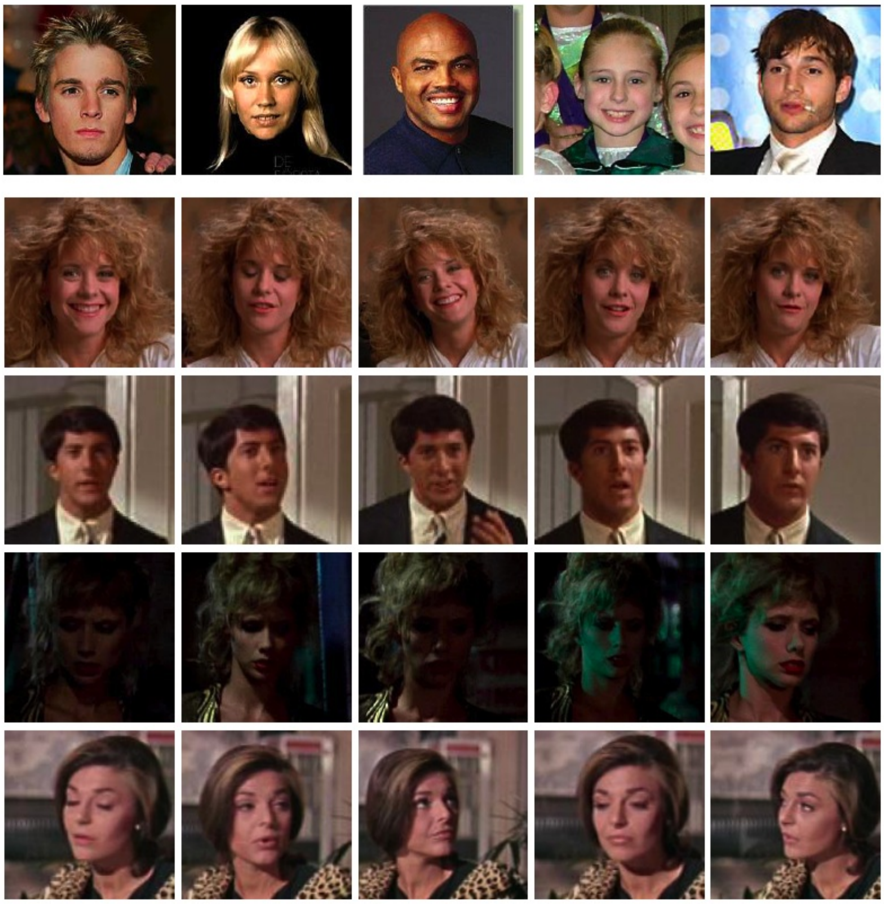

Figure 5. Still images from the FaceTracer database (top row) with versus face tracks from video (four bottom rows). Faces tracks in video contain more variety of expression, lighting, and viewpoint. Examples of different expressions, lighting conditions, and viewpoints within a face track can be automatically associated by tracking and used to improve the face attribute classifier.

correspondence estimation. Instead, it relies on weak geometric properties and combines them with machine learning for efficient cross-view action recognition. The method is validated on three public datasets. It has similar or superior performance compared to related methods and it performs well even in extreme conditions such as when recognizing actions from top views while using side views only for training.

This work resulted in a publication [14].

6.4.5. Data-driven Crowd Analysis in Videos (M. Rodriguez, J.-Y. Audibert, I. Laptev and J. Sivic)

In this work we present a new crowd tracking algorithm that is powered by a large database of crowd videos gathered from the Internet. The algorithm works by generating crowd behavior priors for a specific patch of video after finding similar crowd patch regions in the database. We adhere to the insight that despite the fact that the entire space of possible crowd behaviors is infinite, the space of distinguishable crowd motion patterns may not be all that large. For many individuals in a crowd, we are able to find analogous crowd patches in our database which contain similar patterns of behavior that can effectively act as priors to constrain the difficult task of tracking an individual in a crowd. Our algorithm is data-driven and, unlike some crowd characterization methods, does not require us to have seen the test video beforehand. In our experiments, we demonstrate the ability to track people performing common crowd behaviors, as well as individuals taking part in rare crowd events.

This work is under submission [44].

6.4.6. Track to the future: Spatio-temporal video segmentation with long-range motion cues (J. Lezama, K. Alahari, I. Laptev and J. Sivic)

Video provides rich visual cues such as motion and appearance but also much less explored long-range temporal interactions among objects. We aim to capture such interactions and to construct powerful intermediate-level video representation for subsequent recognition. Motivated by this goal, we seek to obtain spatio-temporal oversegmentation of the video into regions that respect object boundaries and, at the same time, associate object pixels over many video frames. The contributions of this paper are twofold. First, we develop an efficient spatio-temporal video segmentation algorithm, that naturally incorporates long-range motion cues from the past and future frames in the form of clusters of point tracks with coherent motion. Second, we devise a new track clustering cost-function that includes occlusion reasoning, in the form of depth ordering constraints, as well as motion similarity along the tracks. We evaluate the proposed approach on a challenging set of video sequences of office scenes from feature length movies.

This work is under submission [40].

6.5. Machine learning for computer vision

6.5.1. Many-to-Many Graph Matching: a Continuous Relaxation Approach (F. Bach, in collaboration with M. Zaslavskiy and J.-P. Vert, Ecole des Mines de Paris)

Graphs provide an efficient tool for object representation in various computer vision applications. Once graph-based representations are constructed, an important question is how to compare graphs. This problem is often formulated as a graph matching problem where one seeks a mapping between vertices of two graphs which optimally aligns their structure. In the classical formulation of graph matching, only one-to-one correspondences between vertices are considered. However, in many applications, graphs cannot be matched perfectly and it is more interesting to consider many-to-many correspondences where clusters of vertices in one graph are matched to clusters of vertices in the other graph. In this work, we formulate the many-to-many graph matching problem as a discrete optimization problem and propose an approximate algorithm based on a continuous relaxation of the combinatorial problem. We compare favorably our method with other existing methods on several benchmark computer vision datasets.

This project resulted in a publication [48].

6.5.2. Discriminative clustering for image co-segmentation (Armand Joulin, Francis Bach and Jean Ponce)

Purely bottom-up, unsupervised segmentation of a single image into foreground and background regions remains a challenging task for computer vision. Co-segmentation is the problem of simultaneously dividing multiple images into regions (segments) corresponding to different object classes. In this paper, we combine existing tools for bottom-up image segmentation such as normalized cuts, with kernel methods commonly used in object recognition. These two sets of techniques are used within a discriminative clustering framework: the goal is to assign foreground/background labels jointly to all images, so that a supervised classifier trained with these labels leads to maximal separation of the two classes. In practice, we obtain a combinatorial optimization problem which is relaxed to a continuous convex optimization problem, that can itself be solved efficiently for up to dozens of images. We illustrate the proposed method on images with very similar foreground objects, as well as on more challenging problems with objects with higher intra-class variations.

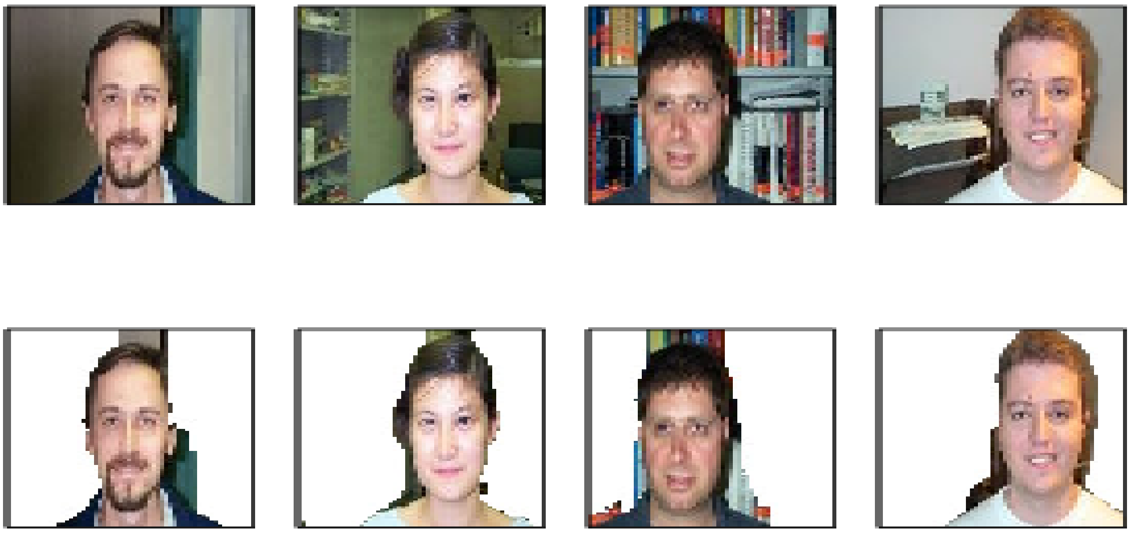

Figure 6. Example of images and their segmentations using our algorithm.

This project illustrated on Figure 6 resulted in a publication [35].

6.5.3. Optimization for Discriminative Latent Class Models (Armand Joulin, Francis Bach and Jean Ponce)

Dimensionality reduction is commonly used in the setting of multi-label supervised classification to control the learning capacity and to provide a meaningful representation of the data. We introduce a simple forward probabilistic model which is a multinomial extension of reduced rank regression, and show that this model provides a probabilistic interpretation of discriminative clustering methods with added benefits in terms of number of hyperparameters and optimization. While the expectation-maximization (EM) algorithm is commonly used to learn these probabilistic models, it usually leads to local maxima because it relies on a non-convex cost function. To avoid this problem, we introduce a local approximation of this cost function, which in turn leads to a quadratic non-convex optimization problem over a product of simplices. In order to maximize quadratic functions, we propose an efficient algorithm based on convex relaxations and low-rank representations of the data, capable of handling large-scale problems. Experiments on text document classification show that the new model outperforms other supervised dimensionality reduction methods, while simulations on unsupervised clustering show that our probabilistic formulation has better properties than existing discriminative clustering methods.

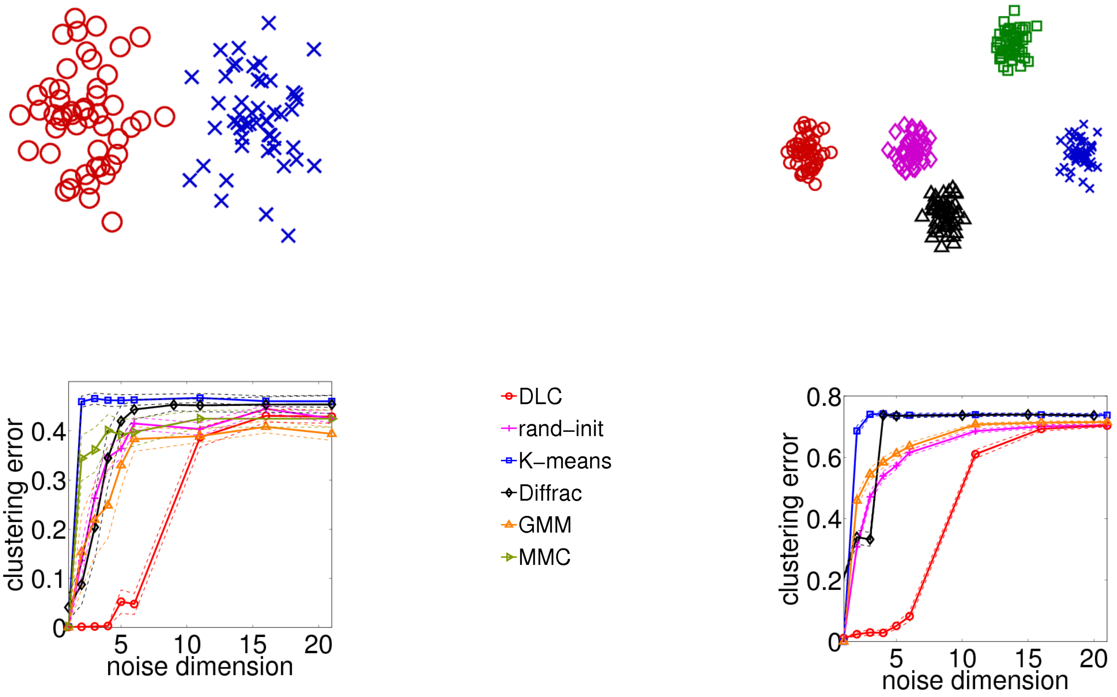

Figure 7. Clustering error when increasing the number of noise dimensions. Our method (DLC) outperform standard clustering methods.

This project illustrated in Figure 7 resulted in a publication [36]

6.5.4. Context-Dependent Kernels for Object Classification (H. Sahbi, J-Y. Audibert, R. Keriven)

Kernels are functions designed in order to capture resemblance between data and they are used in a wide range of machine learning techniques including support vector machines (SVMs). In their standard version, commonly used kernels such as the Gaussian one, show reasonably good performance in many classification and recognition tasks in computer vision, bio-informatics and text processing. In the particular task of object recognition, the main deficiency of standard kernels, such as the convolution one, resides in the lack in capturing the right geometric structure of objects while also being invariant. We focus in this paper on object recognition using a new type of kernel referred to as "context-dependent". Objects, seen as constellations of interest points are matched by minimizing an energy function mixing (1) a fidelity term which measures the quality of feature matching, (2) a neighborhood criterion which captures the object geometry and (3) a regularization term. We will show that the fixed-point of this energy is a context-dependent kernel which is also positive definite. Experiments conducted on object recognition show that when plugging our kernel in SVMs, we clearly outperform SVMs with context-free kernels.

This work is a follow-up to our 2008 CVPR conference paper, and resulted in a journal publication [21].

6.5.5. Online Learning for Matrix Factorization and Sparse Coding (J. Mairal, F.Bach, J. Ponce and G. Sapiro)

Sparse coding—that is, modelling data vectors as sparse linear combinations of basis elements—is widely used in machine learning, neuroscience, signal processing, and statistics. This paper focuses on the large-scale matrix factorization problem that consists of *learning* the basis set, adapting it to specific data. Variations of this problem include dictionary learning in signal processing, non-negative matrix factorization and sparse principal component analysis. In this paper, we propose to address these tasks with a new online optimization algorithm, based on stochastic approximations, which scales up gracefully to large datasets with millions of training samples, and extends naturally to various matrix factorization formulations, making it suitable for a wide range of learning problems. A proof of convergence is presented, along with experiments with natural images and genomic data demonstrating that it leads to state-of-the-art performance in terms of speed and optimization for both small and large datasets.

This project resulted in a publication [18].

6.5.6. Task-Driven Dictionary Learning (Julien Mairal, F.Bach and J.Ponce)

We propose in this paper to unify two different approaches to image and video restoration: On the one hand, learning a basis set (dictionary) adapted to sparse signal descriptions has proven to be very effective in image reconstruction and classification tasks. On the other hand, explicitly exploiting the self-similarities of natural images has led to the successful non-local means approach. We propose simultaneous sparse coding as a framework for combining these two ideas in a natural manner. This is achieved by jointly decomposing groups of similar signals on subsets of the learned dictionary. As a result, the coefficients of the sparse decompositions are more stable and the quality of the reconstructed images is improved. Experimental results in denoising and demosaicking tasks for image and video data, with synthetic and real non uniform noise, show that the proposed method achieves state-of-the-art results, making it possible to effectively restore raw images from digital cameras at a reasonable speed and memory cost.

This project resulted in a publication [57].

6.5.7. PhD Thesis: Représentations parcimonieuses en apprentissage statistique, traitement d'image et vision par ordinateur (Julien Mairal)

We study in this thesis [2] a particular machine learning approach to represent signals that that consists of modelling data as linear combinations of a few elements from a learned dictionary. It can be viewed as an extension of the classical wavelet framework, whose goal is to design such dictionaries (often orthonormal basis) that are adapted to natural signals. An important success of dictionary learning methods has been their ability to model natural image patches and the performance of image denoising algorithms that it has yielded. We address several open questions related to this framework: How to efficiently optimize the dictionary? How can the model be enriched by adding a structure to the dictionary? Can current image processing tools based on this method be further improved? How should one learn the dictionary when it is used for a different task than signal reconstruction? How can it be used for solving computer vision problems? We answer these questions with a multidisciplinarity approach, using tools from statistical machine learning, convex and stochastic optimization, image and signal processing, computer vision, but also optimization on graphs.

6.6. Machine Learning: Algorithms and Learning Theory

6.6.1. Margin adaptive model selection in statistical learning (S. Arlot, joint work with Peter L. Bartlett, University of California at Berkeley)

A classical condition for fast learning rates is the margin condition, first introduced by Mammen and Tsybakov. We tackle in this paper the problem of adaptivity to this condition in the context of model selection, in a general learning framework. Actually, we consider a weaker version of this condition that allows us to take into account that learning within a small model can be much easier than in a large one. Requiring this "strong margin adaptivity" makes the model selection problem more challenging. We first prove, in a very general framework, that some penalization procedures (including local Rademacher complexities) exhibit this adaptivity when the models are nested. Contrary to previous results, this holds with penalties that only depend on the data. Our second main result is that strong margin adaptivity is not always possible when the models are not nested: for every model selection procedure (even a randomized one), there is a problem for which it does not demonstrate strong margin adaptivity [3].

6.6.2. A survey of cross-validation procedures for model selection (S. Arlot, joint work with Alain Celisse, Université Lille 1)

Used to estimate the risk of an estimator or to perform model selection, cross-validation is a widespread strategy because of its simplicity and its apparent universality. Many results exist on the model selection performances of cross-validation procedures. This survey intends to relate these results to the most recent advances of model selection theory, with a particular emphasis on distinguishing empirical statements from rigorous theoretical results. As a conclusion, guidelines are provided for choosing the best cross-validation procedure according to the particular features of the problem at hand [6].

6.6.3. Structured Sparsity-Inducing Norms through Submodular Functions (F. Bach)

The concept of parsimony is central in many scientific domains. In the context of statistics, signal processing or machine learning, it takes the form of variable or feature selection problems, and is commonly used in two situations: First, to make the model or the prediction more interpretable or cheaper to use, i.e., even if the underlying problem does not admit sparse solutions, one looks for the best sparse approximation. Second, sparsity can also be used given prior knowledge that the model should be sparse. In these two situations, reducing parsimony to finding models with low cardinality turns out to be limiting, and structured parsimony has emerged as a fruitful practical extension, with applications to image processing, text processing or bioinformatics. In this work, we investigate more general set-functions than the cardinality, that may incorporate prior knowledge or structural constraints which are common in many applications: namely, we show that for nondecreasing submodular set-functions, the corresponding convex envelope can be obtained from its Lovász extension, a common tool in submodular analysis. This defines a family of polyhedral norms, for which we provide generic algorithmic tools (subgradients and proximal operators) and theoretical results (conditions for support recovery or high-dimensional inference).

This project resulted in a conference paper [24] and a tutorial paper [59].

6.6.4. Online Learning for Latent Dirichlet Allocation (F. Bach, in collaboration with M. Hoffman and D. Blei, Princeton University)

Hierarchical Bayesian modeling has become a mainstay in machine learning and applied statistics. Bayesian models provide a natural way to encode assumptions about observed data, and analysis proceeds by examining the posterior distribution of model parameters and latent variables conditioned on a set of observations. For example, research in probabilistic topic modeling—the application we will focus on in this work—revolves around fitting complex hierarchical Bayesian models to large collections of documents. In this work, we develop an online variational Bayes (VB) algorithm for Latent Dirichlet Allocation (LDA). Online LDA is based on online stochastic optimization with a natural gradient step, which we show converges to a local optimum of the VB objective function. It can handily analyze massive document collections, including those arriving in a stream. We study the performance of online LDA in several ways, including by fitting a 100-topic topic model to 3.3M articles from Wikipedia in a single pass.

This project resulted in a publication [32].

6.6.5. Self-Concordant Analysis for Logistic Regression (F. Bach)

Most of the non-asymptotic theoretical work in regression is carried out for the square loss, where estimators can be obtained through closed-form expressions. In this paper, we use and extend tools from the convex optimization literature, namely self-concordant functions, to provide simple extensions of theoretical results for the square loss to the logistic loss. We apply the extension techniques to logistic regression with regularization by the ℓ_2-norm and regularization by the ℓ_1-norm, showing that new results for binary classification through logistic regression can be easily derived from corresponding results for least-squares regression.

This project resulted in a publication [9].

6.6.6. Low-Rank Optimization on the Cone of Positive Semidefinite Matrices (in collaboration with R. Sepulchre, Université de Liège et P.-A. Absil, Université catholique de Louvain)

Many combinatorial optimization problems can be relaxed into a convex program. These relaxations are mainly introduced as a tool to obtain lower and upper bounds on the problem of interest. The relaxed solutions provide approximate solutions to the original program. Even when the relaxation is convex, computing its solution might be a demanding task in the case of large-scale problems. In fact, some convex relaxations of combinatorial problems consist in expanding the dimension of the search space by optimizing over a symmetric positive semidefinite matrix variable of the size of the original problem. Fortunately, in many cases, the relaxation is tight once its solution is rank one, and it is expected that the convex relaxation, de?ned in terms of a matrix variable that is likely to be very large, presents a low-rank solution. We exploit this property by designing a low-rank nonconvex formulation that allows to provably and efficiently find the global solution of the large convex problem. This is done using second order optimization on properly defined manifolds.

This project resulted in a publication [13]

6.6.7. Structured Sparse Principal Component Analysis (R. Jenatton, G. Obozinski and F. Bach)

We present an extension of sparse PCA, or sparse dictionary learning, where the sparsity patterns of all dictionary elements are structured and constrained to belong to a prespecified set of shapes. This structured sparse PCA is based on a structured regularization recently introduced by Jenatton et al. (2009). While classical sparse priors only deal with cardinality, the regularization we use encodes higher-order information about the data. We propose an efficient and simple optimization procedure to solve this problem. Experiments with two practical tasks, the denoising of sparse structured signals and face recognition, demonstrate the benefits of the proposed structured approach over unstructured approaches.

This work illustrated in Figure 8 has been published in [34].

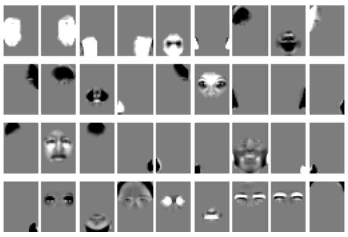

Figure 8. Example of a dictionary learned on a face database.

6.6.8. Proximal Methods for Hierarchical Sparse Coding (R. Jenatton, J. Mairal, G. Obozinski, F. Bach)

Sparse coding consists in representing signals as sparse linear combinations of atoms selected from a dictionary. We consider an extension of this framework where the atoms are further assumed to be embedded in a tree. This is achieved using a recently introduced tree-structured sparse regularization norm, which has proven useful in several applications. This norm leads to regularized problems that are difficult to optimize, and we propose in this paper efficient algorithms for solving them.

This project illustrated on Figure 9 resulted in two publications [33], [12].

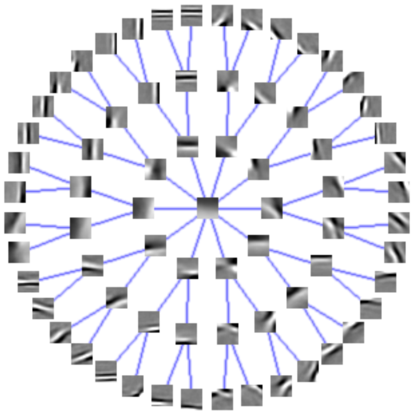

Figure 9. Example of a learned hierarchical dictionary of image patches.

6.6.9. Network Flow Algorithms for Structured Sparsity (J.Mairal, R.Jenatton, G.Obozinski and F.Bach)

We consider a class of learning problems that involve a structured sparsity-inducing norm defined as the sum of ℓ_∞-norms over groups of variables. Whereas a lot of effort has been put in developing fast optimization methods when the groups are disjoint or embedded in a specific hierarchical structure, we address here the case of general overlapping groups. To this end, we show that the corresponding optimization problem is related to network flow optimization. More precisely, the proximal problem associated with the norm we consider is dual to a *quadratic min-cost flow problem*. We propose an efficient procedure which computes its solution exactly in polynomial time. Our algorithm scales up to millions of variables, and opens up a whole new range of applications for structured sparse models. We present several experiments on image and video data, demonstrating the applicability and scalability of our approach for various problems.

This project resulted in two publications [19], [41].

6.6.10. Sequential prediction (J.-Y. Audibert, S. Bubeck, Universitat Autónoma de Barcelona)

We study four classical sequential prediction settings, namely full information, bandit, label efficient and bandit label efficient as well as four different notions of regret: pseudo-regret, expected regret, high probability regret and tracking the best expert regret. We introduce a new forecaster, INF (Implicitly Normalized Forecaster) based on an arbitrary function ψ for which we propose a unified analysis of its pseudo-regret in the four games we consider. In particular, for $\psi(x) = exp(hx) + \gamma$, INF reduces to the classical exponentially weighted average forecaster and our analysis of the pseudo-regret recovers known results while for the expected regret we slightly tighten the bounds. On the other hand with $\psi(x) = (-\eta x)^{-q} + \gamma$, which defines a new forecaster, we are able to remove the extraneous logarithmic factor in the pseudo-regret bounds for bandits games, and thus fill in a long open gap in the characterization of the minimax rate for the pseudo-regret in the bandit game. We also provide high probability bounds depending on the cumulative reward of the optimal action.

This work is an extension of our 2009 COLT conference paper.
It is published in the journal paper [8].

6.6.11. Minimax policy in the stochastic bandit game (J.-Y. Audibert, S. Bubeck)

In the stochastic bandit game, we prove that an appropriate modification of the upper confidence bound policy UCB1 (Auer et al., 2002a) achieves the distribution-free optimal rate while still having a distribution-dependent rate logarithmic in the number of plays. This work is an improved version of our 2009 COLT conference paper, and is published in Section 9 of the journal paper [8].

6.6.12. Best Arm Identification in Multi-Armed Bandits (J.-Y. Audibert, S. Bubeck, R. Munos, SEQUEL team, INRIA Lille)

We consider the problem of finding the best arm in a stochastic multi-armed bandit game. The regret of a forecaster is here defined by the gap between the mean reward of the optimal arm and the mean reward of the ultimately chosen arm. We propose a highly exploring UCB policy and a new algorithm based on successive rejects. We show that these algorithms are essentially optimal since their regret decreases exponentially at a rate which is, up to a logarithmic factor, the best possible. However, while the UCB policy needs the tuning of a parameter depending on the unobservable hardness of the task, the successive rejects policy benefits from being parameter-free, and also independent of the scaling of the rewards. As a by-product of our analysis, we show that identifying the best arm (when it is unique) requires a number of samples of order (up to a $\log(K)$ factor) $\sum_i 1/\Delta_i^2$, where the sum is on the suboptimal arms and Δ_i represents the difference between the mean reward of the best arm and the one of arm i. This generalizes the well-known fact that one needs of order of $1/\Delta^2$ samples to differentiate the means of two distributions with gap Δ.

This work resulted in a COLT conference publication [22] and was integrated in the book chapter [49].

6.6.13. Gaussian process bandit problems (S. Grünewälder, J.-Y. Audibert, M. Opper, J. Shawe-Taylor, from TU Berlin and University College London)

Bandit algorithms are concerned with trading exploration with exploitation where a number of options are available but we can only learn their quality by experimenting with them. We consider the scenario in which the reward distribution for arms is modelled by a Gaussian process and there is no noise in the observed reward. Our main result is to bound the regret experienced by algorithms relative to the a posteriori optimal strategy of playing the best arm throughout based on benign assumptions about the covariance function defining the Gaussian process. We further complement these upper bounds with corresponding lower bounds for particular covariance functions demonstrating that in general there is at most a logarithmic looseness in our upper bounds.

This work resulted in a publication [31].

6.7. Creation of the SIERRA project-team

6.7.1. From WILLOW alone to WILLOW and SIERRA

The WILLOW team officially started in the Spring of 2007. From the start, it was clear that machine learning was a key ingredient to new breakthroughs, and our activities have steadily grown in this area. In three short years, WILLOW has grown into a mature group of about 30 people, and it divides its activities between computer vision, machine learning, and the cross-pollination of the two fields, with video as one of the core research areas. We have been very successful, with many publications in all the major international conferences and leading journals in both areas, but we are a large group with very diverse interests, ranging from camera geometry to statistics, and from image retrieval to bioinformatics applications of structured sparse coding. With the creation of the SIERRA project-team, the core machine learning activities of WILLOW will be transferred to the new group.

The two teams will continue collaborating with each other (they will remain co-located at the INRIA site in central Paris), but they will have a sharper focus on their respective computer vision and machine learning activities.

6.7.2. SIERRA

The SIERRA project-team was created by the INRIA on January 1st 2011 and will be headed by Francis Bach, who received in 2009 a Jr. ERC grant.

Its general academic positioning is described in this section.

Scientific domain.
Machine learning has emerged as its own scientific domain in the last 30 years, providing a good abstraction of many problems and allowing exchanges of best practices between data-oriented scientific fields. Its main research areas are currently probabilistic models, supervised learning, unsupervised learning, reinforcement learning, and statistical learning theory. All of these are represented in SIERRA, but the main goals of the team are mostly related to supervised learning, unsupervised learning, and their mutual interactions. One particularity of the team is the strong focus on optimization (bandit learning and convex optimization) and on parsimony (sparsity-inducing norms and model selection).

Research driven by interdisciplinary collaborations.
Machine learning research can be conducted from two main perspectives: the first one, which has been dominant in the last 30 years, is to design learning algorithms and theories which are as generic as possible, the goal being to make as few assumptions as possible regarding the problems to be solved and to let data speak for themselves. This has led to many interesting methodological developments and successful applications. However, we believe that this strategy has reached its limit for many application domains, such as computer vision, bioinformatics, neuro-imaging, text and audio processing, which leads to the second perspective our team is built on: Research in machine learning theory and algorithms should be driven by interdisciplinary collaborations, so that specific prior knowledge may be properly introduced into the learning process.

Triple objective.

Machine learning is at the intersection of several scientific fields (statistics, applied mathematics and computer science) and is being used in most data-oriented fields (e.g., computer vision, signal processing, natural language processing). Our goal is to contribute to all of these while remaining focused on machine learning. We strive to achieve a triple objective, which we believe is necessary for high impact: rigorous theory, efficient algorithms, and improved performance in applications. In other words we aim to bridge the gap between the theory and practice of machine learning.

7. Contracts and Grants with Industry

7.1. EADS (ENS)

Participants: Jean Ponce, Josef Sivic, Andrew Zisserman.

A. Zisserman's participation in WILLOW has been partially funded by EADS. This has resulted in collaboration efforts via tutorial presentations and discussions with A. Zisserman, J. Sivic and J. Ponce at EADS and ENS. In addition, Marc Sturzel (EADS) is doing a PhD at ENS with Jean Ponce and Andrew Zisserman.

7.2. MSR-INRIA joint lab: Image and video mining for science and humanities (INRIA)

Participants: Jean Ponce, Francis Bach, Andrew Zisserman, Josef Sivic, Ivan Laptev.

This collaborative project, already mentioned several times in this report, brings together the WILLOW and LEAR project-teams with MSR researchers in Cambridge and elsewhere. The concept builds on several ideas articulated in the "2020 Science" report, including the importance of data mining and machine learning in computational science. Rather than focusing only on natural sciences, however, we propose here to expand the breadth of e-science to include humanities and social sciences. The project we propose will focus on fundamental computer science research in computer vision and machine learning, and its application to archaeology, cultural heritage preservation, environmental science, and sociology, and it will be validated by collaborations with researchers and practitioners in these fields. Total budget: 628 KEuros.

7.3. QUAERO (INRIA)

Participant: Ivan Laptev.

QUAERO (AII) is a European collaborative research and development program with the goal of developing multimedia and multi-lingual indexing and management tools for professional and public applications. Quaero consortium involves 24 academic and industrial partners leaded by Technicolor (previously Thomson). Willow participates in work package 9 “Video Processing” and leads work on motion recognition and event recognition tasks. Total funding for WILLOW: 482 KEuros.

7.4. CrowdChecker (ENS)

Participants: Jean-Yves Audibert, Jean Ponce, Josef Sivic, Ivan Laptev.

CrowdChecker (DGA) is a joint project with industrial partner E-vitech. This contract belongs to our video understanding research program. It aims at real-time characterization of a crowd seen from a camera mounted 3 to 10 meters over the ground. It includes segmentation of the crowd, clustering by movement, detection of abnormal behaviors (persons, for instance, crossing the crowd flow, or having unusual speed), tracking people. Several parts of computer vision and machine learning are involved: crowd optical flow estimation, image processing, crowd feature extraction, statistical learning from video database, etc. Total WILLOW funding: 70 KEuros.

7.5. PersonSpace (INRIA)

Participant: Ivan Laptev.

PersonSpace is a CIFRE PhD contract with Technicolor-R&D. The project addresses the problem of human pose estimation and human action recognition in still images. We investigate a subspace spanned by images and videos of people and explore the structure of this subspace to formulate useful constraints for automatic interpretation of person images. Total funding for WILLOW: 15 KEuros.

8. Other Grants and Activities

8.1. Agence Nationale de la Recherche: HFIMBR (INRIA)

Participants: Jean Ponce, Josef Sivic, Oliver Whyte, Andrew Zisserman.

This is a collaborative effort with A. Bartoli (LASMEA Clermont-Ferrand) and N. Holszuch (ARTIS project-team, INRIA Rhône-Alpes).

There is an increasing need for three-dimensional (3D) “content” in entertainment, engineering, and scientific applications. We predict that, for most of these, today’s specialized 3D sensors will eventually be replaced by ordinary, consumer-grade digital cameras equipped with advanced image-based modeling and analysis software. We propose core computer vision and computer graphics research that will enable the development of this software and its application to real-world problems. Concretely, in Willow, we focus on high-fidelity image-based modeling, 3D shape/appearance matching and image/video enhancement and restoration from multiple un-calibrated photographs. The goal is to demonstrate applications of the technology developed in this project to film post production and special effects, and cultural heritage conservation, both pursued via collaborations with external partners. Total funding for WILLOW: 110 KEuros.

8.2. Agence Nationale de la Recherche: DETECT (ENS)

Participants: Sylvain Arlot, Francis Bach, Josef Sivic.

The DETECT project aims at providing new statistical approaches for detection problems in computer vision (in particular, detecting and recognizing human actions in videos) and bioinformatics (e.g., simultaneously segmenting CGH profiles). These problems are mainly of two different statistical nature: multiple change-point detection (i.e., partitioning a sequence of observations into homogeneous contiguous segments) and multiple tests (i.e., controlling a priori the number of false positives among a large number of tests run simultaneously).

This is a collaborative effort with A. Celisse (University Lille 1), T. Mary-Huard (AgroParisTech), E. Roquain and F. Villers (Univeristy Paris 6), in addition to S. Arlot, F. Bach and J. Sivic from Willow.

S. Arlot is the leader of this ANR “Young researchers” project. The total funding is 70000 Euros.

8.3. Agence Nationale de la Recherche: MGA (INRIA/ENPC)

Participants: Jean-Yves Audibert, Francis Bach, Olivier Duchenne, Julien Mairal, Jean Ponce, Andrew Zisserman.

Probabilistic graphical models, also known as Bayesian Networks, provide a very flexible and powerful framework for capturing statistical dependencies in complex, multivariate data. They enable the building of large global probabilistic models for complex phenomena out of smaller and more tractable local models. The objectives of this project are to advance the methodological state of the art of probabilistic modeling research, while applying the newly developed techniques to computer vision, text processing and bio-informatics. F. Bach is the coordinator of this ANR “projet blanc” in machine learning, that focuses on graphical models and their applications. The total funding is 200 KEuros, with 100KEuros for Willow including (50KEuros for INRIA and 50KEuros for ENPC).

8.4. Agence Nationale de la Recherche: Triangles (ENS)

Participant: Jean Ponce.

This is a collaborative effort with O. Devillers (INRIA project-team GEOMETRICA), Raphaëlle Chaine (University of Lyon), and J. Ponce and E. Colin de Verdière (ENS).

This project is dedicated to the design of computational geometry methods for constructing triangulation in non-Euclidean spaces. Total funding for WILLOW: 5000 Euros.

8.5. European Research Council (ERC) Advanced Grant

Participants: Jean Ponce, Ivan Laptev, Josef Sivic.

WILLOW will be funded in part from 2011 to 2015 by the ERC Advanced Grant "VideoWorld" awarded to Jean Ponce by the European Research Council.

This project is concerned with the automated computer analysis of video streams: Digital video is everywhere, at home, at work, and on the Internet. Yet, effective technology for organizing, retrieving, improving, and editing its content is nowhere to be found. Models for video content, interpretation and manipulation inherited from still imagery are obsolete, and new ones must be invented. With a new convergence between computer vision, machine learning, and signal processing, the time is right for such an endeavor. Concretely, we will develop novel spatio-temporal models of video content learned from training data and capturing both the local appearance and nonrigid motion of the elements—persons and their surroundings—that make up a dynamic scene. We will also develop formal models of the video interpretation process that leave behind the architectures inherited from the world of still images to capture the complex interactions between these elements, yet can be learned effectively despite the sparse annotations typical of video understanding scenarios. Finally, we will propose a unified model for video restoration and editing that builds on recent advances in sparse coding and dictionary learning, and will allow for unprecedented control of the video stream. This project addresses fundamental research issues, but its results are expected to serve as a basis for groundbreaking technological advances for applications as varied as film post-production, video archival, and smart camera phones.

8.6. European Research Council (ERC) Starting Investigator Researcher grant

Participants: Francis Bach, Guillaume Obozinski.

The new SIERRA team-project will be funded in part by a grant from the European Research Council (ERC) and coordinated by Francis Bach. The goals of the project are to explore sparse structured methods for machine learning, with applications in computer vision and audio processing.

9. Dissemination

9.1. Animation of the scientific community

+ Conference and workshop organization

  - S. Arlot, Organizer of a session “Model selection” inside Journées MAS 2010 (SMAI), Bordeaux. http://www.math.u-bordeaux1.fr/MAS10/session3G5.html
  - J.-Y. Audibert, Co-organizer, Foundations and New Trends of PAC Bayesian Learning Workshop, London, Great Britain, 2010
  - I. Laptev, J. Ponce and J. Sivic, Organizers, MSR-INRIA workshop on computer vision and machine learning, http://www.di.ens.fr/willow/events/msr-inria2010/
  - A. Zisserman, Co-organizer of the PASCAL Visual Object Classes Challenge 2010 (VOC2010). http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2010/
  - A. Zisserman, Co-organizer of the PASCAL VOC 2010 workshop at ECCV 2010 http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2010/workshop/index.html
  - A. Zisserman, Organizer of the Oxford Brookes/KTH/VGG/Willow Workshop held in Oxford in July 2010, see http://www.robots.ox.ac.uk/~vgg/workshops/oxford10/programme.html

+ Editorial Boards

  - Journal of Machine Learning Research, Action Editor (F. Bach).
  - International Journal of Computer Vision (I. Laptev, J. Ponce, J. Sivic and A. Zisserman).
  - Image and Vision Computing Journal (I. Laptev).
  - Fondations and Trends in Computer Graphics and Vision (J. Ponce).
  - SIAM Journal on Imaging Sciences (J. Ponce, F. Bach)
  - IEEE Transactions Pattern Analysis and Machine Intelligence (F. Bach)

+ Area Chairs

  - IEEE Conference on Computer Vision and Pattern Recognition, 2010 (I. Laptev).
  - IEEE International Conference on Automatic Face and Gesture Recognition, 2011 (I. Laptev).
  - IEEE International Conference on Computer Vision, 2011 (I. Laptev and J. Sivic).
  - International Conference on Artificial Intelligence and Statistics, 2010 (F.Bach).
  - European Conference on Computer Vision, 2010 (F. Bach).
  - International Conference on Machine Learning, 2010 (F. Bach).

+ Program Committees

  - IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010 (J. Sivic, F. Bach, T. Cour, I. Laptev, A. Zisserman, J. Mairal).
  - European Conference on Computer Vision (ECCV), 2010 (A. Zisserman, J. Mairal, J. Sivic, I. Laptev, R. Jenatton).
  - Reviewer for the ACM International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH), 2010 (J. Sivic).
  - Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP), 2010 (A. Zisserman).
  - International Conference on Neural Information Processing Systems (NIPS), 2010 (J. Mairal, G. Obozinski, J. Sivic, R. Jenatton, I.Laptev,J.-Y. Audibert).
  - International Conference in Machine Learning (ICML), 2010 (A. Zisserman, J. Mairal, G. Obozinski, J. Sivic, R. Jenatton).
  - International Conference on Artificial Intelligence and Statistics (AISTATS), 2010 (J. Mairal, G. Obozinski, R. Jenatton).
  - International Conference on Robotics and Automation (ICRA), 2010 (A. Zisserman).
  - British Machine Vision Conference (BMVC), 2010 (I. Laptev).
  - Conférence francophone sur lapprentissage automatique (CAp), 2010 (J.-Y. Audibert)

+ PhD and HDR thesis committee:

  - Sarah Filippi, Télécom ParisTech, 2010 (J.-Y. Audibert)
  - Anne-Marie Tousch, Ecole des Ponts ParisTech, 2010 (J.-Y. Audibert)
  - Philippe Rolet, Université d'Orsay, 2010 (J.-Y. Audibert)
  - Antoine Salomon, Université Paris 13, 2010 (J.-Y. Audibert)
  - Xinhua Zhang, Australian National University, 2010 (F. Bach)
  - Nataliya Sokolovska, Telecom Paristech, 2010 (F. Bach)
  - Vincent Michel, Université Paris-Sud, 2010 (F. Bach)
  - Joaquin Zepeda, Université de Rennes 1, 2010 (F. Bach)
  - Alexander Kläser, INRIA Grenoble, 2010 (I. Laptev)
  - Alonso A Patron-Perez, University of Oxford, 2010 (I. Laptev)
  - Jérome Courchay, Université Paris-Est, 2010 (J. Ponce)
  - Manuel Jesus Marin Jimenez, Spain, 2010, (A. Zisserman)
  - Mark Cummins, Oxford, 2010 (A. Zisserman)

+ Prizes:

  - INRIA Prime d'excellence scientifique (I. Laptev, J. Sivic)
  - Outstanding review awards at ECCV 2010 (I. Laptev)
  - S. Arlot, Cours Peccot 2011, Collège de France. http://www.di.ens.fr/~arlot/peccot.htm
  - Best paper award for Klser, A., Marszalek, M., Schmid, C. and Zisserman, A. Human Focused Action Localization in Video International Workshop on Sign, Gesture, Activity (2010) [38]
  - Best industrial paper prize for Patron-Perez, A., Marszalek, M., Reid, I. and Zisserman, A. High Five: Recognising Human Interactions in TV Shows British Machine Vision Conference (2010) [42]

+ Other:

- S. Arlot is member of the board for the entrance exam in École Normale Supérieure (mathematics, voie B/L)
- J.-Y. Audibert, F. Bach, Co-organizers of the biweekly seminar “Statistical Machine Learning in Paris" (http://sites.google.com/site/smileinparis/home)

9.2. Teaching

- S. Arlot and Francis Bach, “Statistical learning”, Master 2 “Probabilités et Statistiques”, Université Paris-Sud, 24h.
- J.-Y. Audibert, “Machine Learning and applications”, Ecole des Ponts ParisTech, 2nd year, 30h.
- J.-Y. Audibert, “Introduction to Machine Learning”, MVA, Ecole Normale Supérieure de Cachan, 20h.
- J. Mairal, Co-organizer of the tutorial Sparse Coding and Dictionary Learning for Image Analysis, at the conference CVPR 2010, San Francisco.
- J. Mairal, Course given at the summer schools CVML 2010, Grenoble, and ERMITES 2010, Hyéres.
- F. Bach and G.Obozinski, “Probabilistic graphical models”, MVA, Ecole Normale Supérieure de Cachan, 30h.
- G. Obozinski, “Introduction to probabilistic graphical models", Enseignement Spécialisé “Apprentissage artificiel", Ecole des Mines de Paris, 4h.
- G. Obozinski, Tutorial on sparsity, summer school “Sparsity in Image and Signal Analysis", Hólar, Iceland, August 15 - 20, 3h.
- F. Bach and G. Obozinski, Tutorial“Sparse methods for machine learning: Theory and algorithms", European Conference on Machine Learning, Barcelona, September 20th, 3h.
- J. Ponce, "Introduction to computer vision", Ecole normale supérieure, 3h.
- M. Pocchiola and J. Ponce, "Geometric bases of computer science, Ecole normale supérieure, 3h.
- I. Laptev, J. Ponce and J. Sivic (together with C. Schmid (INRIA Grenoble)), “Object recognition and computer vision”, Ecole normale supérieure, and MVA, Ecole normale supérieure de Cachan, 36h.
- I. Laptev (together with G. Mori) co-organized the tutorial on Statistical and Structural Recognition of Human Actions at ECCV 2010, Heraklion, Greece.
- I. Laptev was a speaker at AERFAI Summer School 2010, Benicàssim, Spain, June 2010.
- A. Zisserman, Optimization lectures (Michaelmas 2010), University of Oxford, http://www.robots.ox.ac.uk/~az/lectures/b1/index.html
- A. Zisserman, Machine Learning (Hilary Term 2010), University of Oxford http://www.robots.ox.ac.uk/~az/lectures/ml/index.html
- A. Zisserman, Lab on Information Engineering, University of Oxford
- A. Zisserman, Teaching on two topics at the INRIA Visual Recognition and Machine Learning Summer School http://www.di.ens.fr/willow/events/cvml2010/

9.3. INRIA Visual Recognition and Machine Learning Summer School 2010

http://www.di.ens.fr/willow/events/cvml2010/

I. Laptev and J. Sivic (together with C. Schmid (INRIA Grenoble)) co-organized a one week summer school on Visual Recognition and Machine Learning. The summer school, hosted by INRIA Grenoble, attracted 137 participants from 26 countries (49% France / 35% Europe / 8% Asia and Middle East / 5% North America / 2% South America / 1% Australia), which included Master students, PhD students as well as Post-docs and researchers. The summer school provided an overview of the state of the art in visual recognition and machine learning. Lectures were given by 12 speakers (2 USA, 2 UK, 1 Austria, 7 INRIA / ENS), which included top international experts in the area of visual recognition (D. Forsyth, UIUC, USA; M. Hebert, CMU, USA; A. Zisserman, Oxford, UK / WILLOW). Lectures were complemented by practical sessions to provide participants with hands-on experience with the discussed material. In addition, a poster session was organized for participants to present their current research.

A similar summer school is currently in preparation for 2011 to be hosted by Ecole Normale Supérieure in Paris.

9.4. Invited presentations

- S. Arlot, Study of the Yoccoz-Birkeland model, Seminar, Università degli Studi di Siena, Dipartimento di Ingegneria dell'Informazione, November 2009.
- S. Arlot, Resampling-based estimation of the accuracy of satellite ephemerides, Seminar, Scuola Normale Superiore di Pisa, November 2009.
- S. Arlot, Calibration automatique d'estimateurs linéaires à l'aide de pénalités minimales, Séminaire Probabilités et Statistiques, Université Lille 1, January 2010.
- S. Arlot, Data-driven penalties for optimal calibration of learning algorithms, Workshop "Challenging problems in statistical learning", Paris, January 2010.
- S. Arlot, Data-driven penalties for linear estimator selection, Workshop "Modern Nonparametric Statistics: Going Beyond Asymptotic Minimax", MFO, Oberwolfach, Germany, March 2010.
- S. Arlot, Margin adaptive model selection in statistical learning, Séminaire de Statistiques du CREST, Paris, October 2010.
- S. Arlot, Resampling-based confidence regions and multiple tests, JSTAR 2010: "Statistics in High-dimension", Rennes, October 2010.
- J.-Y. Audibert, Risk bounds for linear regression, Berlin, 2010
- J.-Y. Audibert, Concentration inequalities, Theory of Randomized Search Heuristics workshop, keynote speech, Paris, 2010
- J.-Y. Audibert, Robust estimation in linear least squares regression, IHP, Paris, 2010
- J.-Y. Audibert, PAC-Bayesian bounds and aggregation, Foundations and New Trends of PAC Bayesian Learning, London, Great Britain, 2010
- J.-Y. Audibert, Robust estimation in linear least squares regression, Fréjus, 2010
- F. Bach, Statistics Seminar, ETH Zurich, 2010
- F. Bach, One-day course on sparse methods, Oxford University, 2010
- F. Bach, New-York University, 2010
- F. Bach, Princeton University, 2010
- F. Bach, Université Catholique de Louvain-la-Neuve, 2010
- F. Bach, invited talk, Conference BENELEARN, Leuven, 2010
- F. Bach, One-day course on sparse methods, Microsoft Research Asia, Beijing, 2010
- F. Bach, invited talk, Pao-Lu Hsu Conference, Xian, China, 2010
- F. Bach, invited talk, GDR calcul scientifique, Lyon, 2010

- F. Bach, invited talks, NIPS workshops, 2010
- I. Laptev, GDR-ISIS scientific meeting, Paris, France, December 2010.
- I. Laptev, University Luxembourg, Data Mining Applications, Luxembourg, November 2010.
- I. Laptev, ECCV2010 Workshop on Sign, Gesture and Activity, Grece, September 2010.
- I. Laptev, ICPR2010 Workshop on Analysis and Evaluation of Large-Scale Multimedia Collections, Istanbul, Turkey, August 2010.
- I. Laptev, ICPR2010 Workshop on Human Behaviour Understanding, Istanbul, Turkey, August 2010.
- I. Laptev, Joint Oxford-KTH-INRIA workshop, Oxford, UK, July 2010.
- I. Laptev, International Workshop on Frontiers of Activity Recognition, Los Angeles, USA, June 2010.
- J. Mairal, Colloque STATIM 2010, Evry.
- J. Mairal, Symposium "Statistical Models for Images", Journées de la Statistique, Luminy (http://jds2010.univmed.fr/).
- J. Mairal, Seminar of the INRIA LEAR project-team, Grenoble.
- J. Mairal, Oxford Brookes/KTH/VGG/Willow Workshop, Oxford.
- J. Mairal, Joint MSR-INRIA workshop, Paris.
- G. Obozinski, Statistics Laboratory, University of Cambridge, UK, February 2010.
- G. Obozinski, Statistics Department, University of Yale, CT, February 2010.
- G. Obozinski, Information Systems Group, Stern School of Business, NYU, NY, February 2010.
- G. Obozinski, Statistics Department, University of Pennsylvania, PA, February 2010.
- G. Obozinski, Statistics Department, University of Chicago, IL, February 2010.
- G. Obozinski, Research Seminar, Google, Zürich, Switzerland, February 2010.
- G. Obozinski, Group Lasso Extensions and Structured Sparsity for Dictionary Learning, External Seminar, Gatsby Computational Neuroscience Unit, University College London, UK, February 2010.
- G. Obozinski, Statistics Seminar, Département de Mathématiques Appliquées (MAP5), Université Paris Descartes, Paris, February 2010.
- G. Obozinski, Sustain workshop Sparse structures: statistical theory and practice, University of Bristol, UK, June 2010.
- G. Obozinski, Oxford Brookes/KTH/VGG/Willow Workshop, University of Oxford, UK, July 2010.
- G. Obozinski, Workshop on Inverse Problems in Data Driven Modeling, Johann Radon Institute for Computational and Applied Mathematics (RICAM), Linz, Austria, July 2010.
- G. Obozinski, Journées STAR 2010 "Statistique en grande dimension", Laboratoire de Statistiques, Composante Rennes 2 de l' Institut Mathématiques de Rennes, October 2010.
- G. Obozinski, Groupe de Recherche ISIS "Apprentissage et Parcimonie", November 2010.
- G.Obozinski, NIPS workshop: "New Directions in Multiple Kernel Learning", December 2010.
- J. Ponce, Keynote speaker, British Machine Vision Conference, Aberystwyth, Wales.
- J. Ponce, Distinguished speaker, Dept. of Computer Science, University of Delaware.
- J. Ponce, Janelia Conference on What Can Computer Vision Do for Neuroscience and Vice Versa, VA.
- J. Ponce, Department of Computer Science, New York University, New York City.
- J. Ponce, Laboratoire d'informatique Gaspard-Monge, Paris

- J. Ponce, Laboratoire Jacques-Louis Lions, Paris.
- J. Sivic, Joint Oxford-KTH-INRIA workshop, Oxford, UK, July 2010.
- J. Sivic, New York University, USA, September 2010.
- J. Sivic, DARPA/ARO Workshop on Interactive query refinement for image/video retrieval, September 2010, Columbia University, USA.
- A. Zisserman, Willow MSR-INRIA workshop, January 2010
- A. Zisserman, Keynote speaker at DAGM 2010 http://www.dagm2010.org/index.html

10. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] J.-Y. AUDIBERT. *PAC-Bayesian aggregation and multi-armed bandits*, Université Paris Est, 2010, Habilitation à Diriger des Recherches, http://arxiv.org/abs/1011.3396.

[2] J. MAIRAL. *Représentations parcimonieuses en apprentissage statistique, traitement d'image et vision par ordinateur*, Ecole Normale Supérieure de Cachan, 2010, http://www.di.ens.fr/~mairal/resources/pdf/phd_thesis.pdf.

Articles in International Peer-Reviewed Journal

[3] S. ARLOT, P. L. BARTLETT. *Margin adaptive model selection in statistical learning*, in "Bernoulli", 2010, Accepted. arXiv:0804.2937, http://fr.arxiv.org/PS_cache/arxiv/pdf/0804/0804.2937v2.pdf.

[4] S. ARLOT, G. BLANCHARD, E. ROQUAIN. *Some non-asymptotic results on resampling in high dimension, I: Confidence regions, II: Multiple tests*, in "The Annals of Statistics", 2010, vol. 38, n^{o} 1, p. 51-99, http://hal.inria.fr/hal-00194145/en.

[5] S. ARLOT, G. BLANCHARD, E. ROQUAIN. *Some non-asymptotic results on resampling in high dimension, II: Multiple tests*, in "The Annals of Statistics", 2010, vol. 38, n^{o} 1, p. 83-99, http://hal.inria.fr/hal-00194145/en.

[6] S. ARLOT, A. CELISSE. *A survey of cross-validation procedures for model selection*, in "Statist. Surv.", 2010, vol. 4, p. 40–79 [*DOI :* 10.1214/09-SS054], http://www.di.ens.fr/willow/pdfs/2010_Arlot_Celisse_SS.pdf.

[7] S. ARLOT, A. CELISSE. *Segmentation of the mean of heteroscedastic data via cross-validation*, in "Statistics and Computing", 2010, p. 1–20, http://arxiv.org/pdf/0902.3977v2.pdf.

[8] J.-Y. AUDIBERT, S. BUBECK. *Regret Bounds and Minimax Policies under Partial Monitoring*, in "Journal of Machine Learning Research", Oct 2010, vol. 11, p. 2635-2686,, http://imagine.enpc.fr/publications/papers/JMLR10.pdf.

[9] F. BACH. *Self-Concordant Analysis for Logistic Regression*, in "Electronic Journal of Statistics", 2010, vol. 4, p. 384–414, http://www.di.ens.fr/willow/pdfscurrent/bach_ejs_self_concordance.pdf.

[10] O. DUCHENNE, F. BACH, INSO. KWEON, J. PONCE. *A Tensor-Based Algorithm for High-Order Graph Matching*, in "PAMI", 2011, to appear.

[11] Y. Furukawa, J. Ponce. *Accurate, Dense, and Robust Multi-View Stereopsis*, in "IEEE Trans. Patt. Anal. Mach. Intell.", 2010, vol. 32, n^o 8.

[12] R. Jenatton, J. Mairal, G. Obozinski, F. Bach. *Proximal Methods for Hierarchical Sparse Coding*, in "Journal Machine Learning Research", 2010, to appear.

[13] M. Journée, F. Bach, P.-A. Absil, R. Sepulchre. *Low-rank optimization on the cone of positive semidefinite matrices*, in "SIAM Journal on Optimization", 2010, vol. 20, n^o 5, p. 2327–2351.

[14] I. Junejo, E. Dexter, I. Laptev, P. Pérez. *View-independent action recognition from temporal self-similarities*, in "IEEE Trans. Patt. Anal. Mach. Intell.", 2010, vol. in press, http://www.irisa.fr/vista/Papers/2010_pami_junejo.pdf.

[15] B. Kaneva, J. Sivic, A. Torralba, S. Avidan, William T. Freeman. *Infinite Images: Creating and Exploring a Large Photorealistic Virtual Space*, in "Proceedings of the IEEE", 2010, vol. 98, n^o 8, p. 1391–1407, http://www.di.ens.fr/willow/pdfscurrent/kaneva09b.pdf.

[16] H. Kong, J.-Y. Audibert, J. Ponce. *General road detection from a single image*, in "Image Processing, IEEE Transactions on", 2010, vol. 19, n^o 8, p. 2211–2220, http://imagine.enpc.fr/publications/papers/TIP10a.pdf.

[17] H. Kong, Jean-Yves. Audibert, J. Ponce. *Detecting abandoned objects with a moving camera*, in "Image Processing, IEEE Transactions on", 2010, vol. 19, n^o 8, p. 2201–2210, http://imagine.enpc.fr/publications/papers/TIP10b.pdf.

[18] J. Mairal, F. Bach, J. Ponce, G. Sapiro. *Online Learning for Matrix Factorization and Sparse Coding*, in "Journal of Machine Learning Research", January 2010, vol. 11, n^o 1, p. 19–60, http://jmlr.csail.mit.edu/papers/volume11/mairal10a/mairal10a.pdf, http://hal.inria.fr/inria-00408716/en.

[19] J. Mairal, R. Jenatton, G. Obozinski, F. Bach. *Network Flow Algorithms for Structured Sparsity*, in "Journal Machine Learning Research", 2010, to appear.

[20] J. Philbin, J. Sivic, A. Zisserman. *Geometric Latent Dirichlet Allocation on a Matching Graph for Large-scale Image Datasets*, in "International Journal of Computer Vision", 2010, http://dx.doi.org/10.1007/s11263-010-0363-5.

[21] H. Sahbi, J.-Y. Audibert, R. Keriven. *Context-Dependent Kernels for Object Classification*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", November 2010 [*DOI :* 10.1109/TPAMI.2010.198], http://imagine.enpc.fr/publications/papers/PAMI10.pdf.

International Peer-Reviewed Conference/Proceedings

[22] J.-Y. Audibert, S. Bubeck, R. Munos. *Best Arm Identification in Multi-Armed Bandits*, in "Proceedings of the 23th annual conference on Computational Learning Theory (COLT)", 2010, http://certis.enpc.fr/~audibert/Mes%20articles/COLT10.pdf.

[23] F. Bach. *Structured sparsity-inducing norms through submodular functions*, in "NIPS 2010 : Twenty-Fourth Annual Conference on Neural Information Processing Systems", Canada Vancouver, 2010, p. 118–126, http://hal.inria.fr/hal-00511310/en.

[24] F. BACH. *Structured sparsity-inducing norms through submodular functions*, in "Adv. Neural Info. Proc. Systems", 2010, http://books.nips.cc/papers/files/nips23/NIPS2010_0875.pdf.

[25] G. BATOG, X. GOAOC, J. PONCE. *Admissible linear map models of linear cameras*, in "Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on", IEEE, 2010, p. 1578–1585, http://perso.crans.org/batog/cameras.pdf.

[26] G. BATOG, X. GOAOC, J. PONCE. *Admissible Linear Map Models of Linear Cameras*, in "23rd IEEE Conference on Computer Vision and Pattern Recognition - CVPR 2010", United States San Francisco, IEEE, June 2010, p. 1578 - 1585 [*DOI :* 10.1109/CVPR.2010.5539784], http://hal.inria.fr/inria-00517899/en.

[27] Y-LAN. BOUREAU, F. BACH, Y. LECUN, J. PONCE. *Learning Mid-Level Features for Recognition*, in "Proc. International Conference on Computer Vision and Pattern Recognition (CVPR'10)", IEEE, 2010, http://www.di.ens.fr/willow/pdfs/cvpr10c.pdf.

[28] Y-LAN. BOUREAU, J. PONCE, Y. LECUN. *A theoretical analysis of feature pooling in vision algorithms*, in "Proc. International Conference on Machine learning (ICML'10)", 2010, http://www.di.ens.fr/willow/pdfs/icml2010b.pdf.

[29] N. CHERNIAVSKY, I. LAPTEV, J. SIVIC, A. ZISSERMAN. *Semi-supervised learning of facial attributes in video*, in "The first international workshop on parts and attributes (in conjunction with ECCV 2010)", 2010, http://www.di.ens.fr/willow/pdfs/cherniavsky10.pdf.

[30] V. DELAITRE, I. LAPTEV, J. SIVIC. *Recognizing human actions in still images: a study of bag-of-features and part-based representations*, in "Proceedings of the British Machine Vision Conference", 2010, http://www.di.ens.fr/willow/pdfs/delaitre10.pdf.

[31] S. GRÜNEWÄLDER, J.-Y. AUDIBERT, M. OPPER, J. SHAWE-TAYLOR. *Regret bounds for Gaussian process bandit problems*, in "Proceedings of the 14th International Conference on Artificial Intelligence and Statistics", Chia (Italy), May 2010, http://imagine.enpc.fr/publications/papers/AISTATS10.pdf.

[32] M. HOFFMAN, F. BACH, D. BLEI. *Online Learning for Latent Dirichlet Allocation*, in "Adv. Neural Info. Proc. Systems", 2010, http://books.nips.cc/papers/files/nips23/NIPS2010_1291.pdf.

[33] R. JENATTON, J. MAIRAL, G. OBOZINSKI, F. BACH. *Proximal Methods for Sparse Hierarchical Dictionary Learning*, in "Proceedings of the International Conference on Machine Learning (ICML)", 2010, http://www.icml2010.org/papers/416.pdf.

[34] R. JENATTON, G. OBOZINSKI, F. BACH. *Structured sparse principal component analysis*, in "International Conference on Artificial Intelligence and Statistics (AISTATS)", 2010, http://jmlr.csail.mit.edu/proceedings/papers/v9/jenatton10a/jenatton10a.pdf.

[35] A. JOULIN, F. BACH, J. PONCE. *Discriminative Clustering for Image Co-segmentation*, in "Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)", 2010, http://www.di.ens.fr/~fbach/cosegmentation_cvpr2010.pdf.

[36] A. JOULIN, F. BACH, J. PONCE. *Efficient Optimization for Discriminative Latent Class Models*, in "Advances in Neural Information Processing Systems (NIPS)", 2010, http://books.nips.cc/papers/files/nips23/NIPS2010_1104.pdf.

[37] B. Kaneva, J. Sivic, A. Torralba, S. Avidan, William T. Freeman. *Matching and Predicting Street Level Images*, in "ECCV 2010 Workshop on Vision for Cognitive Tasks", 2010, http://www.di.ens.fr/willow/pdfs/kaneva10a.pdf.

[38] A. Kläser, M. Marszałek, C. Schmid, A. Zisserman. *Human Focused Action Localization in Video*, in "International Workshop on Sign, Gesture, Activity", 2010, http://lear.inrialpes.fr/people/klaeser/research_action_localization.

[39] J. Knopp, J. Sivic, T. Pajdla. *Avoiding confusing features in place recognition*, in "Proceedings of the European Conference on Computer Vision", September 2010, http://www.di.ens.fr/willow/pdfs/knopp10.pdf.

[40] J. Lezama, K. Alahari, I. Laptev, J. Sivic. *Track to the future: Spatio-temporal video segmentation with long-range motion cues*, in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition", 2011, to appear.

[41] J. Mairal, R. Jenatton, G. Obozinski, F. Bach. *Network Flow Algorithms for Structured Sparsity*, in "Advances in Neural Information Processing Systems", 2010, http://books.nips.cc/papers/files/nips23/NIPS2010_1040.pdf.

[42] A. Patron-Perez, M. Marszałek, A. Zisserman, I. D. Reid. *High Five: Recognising Human Interactions in TV Shows*, in "British Machine Vision Conference", 2010, http://www.robots.ox.ac.uk/ActiveVision/Publications/patron_etal_bmvc2010/patron_etal_bmvc2010.html.

[43] J. Philbin, M. Isard, J. Sivic, A. Zisserman. *Descriptor learning for efficient retrieval*, in "Proceedings of the European Conference on Computer Vision", September 2010, http://www.di.ens.fr/willow/pdfs/philbin10b.pdf.

[44] M. Rodriguez, Jean-Yves. Audibert, I. Laptev, J. Sivic. *Data-driven Crowd Analysis in Videos*, in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition", 2011, to appear.

[45] A. Torii, T. Pajdla, J. Sivic. *Read between the views: image-based localization by matching linear combinations of views*, in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition", 2011, to appear.

[46] M. Muneeb. Ullah, S. Parizi, I. Laptev. *Improving bag-of-features action recognition with non-local cues*, in "Proceedings of the British Machine Vision Conference", 2010, http://www.di.ens.fr/willow/pdfscurrent/paper95.pdf.

[47] O. Whyte, J. Sivic, A. Zisserman, J. Ponce. *Non-uniform Deblurring for Shaken Images*, in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition", 2010, http://www.di.ens.fr/willow/pdfs/cvpr10d.pdf.

[48] M. Zaslavskiy, F. Bach, Jean-Philippe. Vert. *Many-to-Many Graph Matching: a Continuous Relaxation Approach*, in "Proc. Europ. Conf. on Machine Learning", 2010, http://www.springerlink.com/content/k562m31730823323/.

Scientific Books (or Scientific Book chapters)

[49] S. Bubeck, J.-Y. Audibert, R. Munos. *Bandit view on noisy optimization*, in "Optimization for Machine Learning", S. Sra, S. Nowozin, S. Wright (editors), MIT Press, 2010, http://certis.enpc.fr/~audibert/Mes%20articles/book_chapter.pdf.

Research Reports

[50] J.-Y. Audibert, O. Catoni. *Risk bounds in linear regression through PAC-Bayesian truncation*, HAL, 2010, 78 pages, http://hal.inria.fr/hal-00360268/en.

[51] J.-Y. Audibert, O. Catoni. *Robust linear least squares regression*, HAL, 2010, 48 pages, http://hal.inria.fr/hal-00522534/en.

[52] J.-Y. Audibert, O. Catoni. *Robust linear regression through PAC-Bayesian truncation*, HAL, 2010, http://hal.inria.fr/hal-00522536/en.

[53] F. Bach. *Shaping Level Sets with Submodular Functions*, HAL, 2010, http://hal.inria.fr/hal-00542949/en.

[54] R. Jenatton, J.-Y. Audibert, F. Bach. *Structured Variable Selection with Sparsity-Inducing Norms*, INRIA, 2010, http://hal.inria.fr/inria-00377732/en.

[55] R. Jenatton, J. Mairal, G. Obozinski, F. Bach. *Proximal Methods for Hierarchical Sparse Coding*, INRIA, 2010, http://hal.inria.fr/inria-00516723/en.

[56] A. Kläser, M. Marszałek, I. Laptev, C. Schmid. *Will person detection help bag-of-features action recognition?*, INRIA, September 2010, n^o RR-7373, http://hal.inria.fr/inria-00514828/en.

[57] J. Mairal, F. Bach, J. Ponce. *Task-Driven Dictionary Learning*, INRIA, September 2010, n^o RR-7400, http://hal.inria.fr/inria-00521534/en.

Other Publications

[58] S. Arlot. *Sélection de modèles*, August 2010, Type : Conference digest, http://hal.inria.fr/inria-00496738/en.

[59] F. Bach. *Convex Analysis and Optimization with Submodular Functions: a Tutorial*, 2010, Tutorial, http://hal.inria.fr/hal-00527714/en.