Fortschritt-Berichte VDI

Reihe 10

Informatik/
Kommunikation

Nr. 859

M.Sc. Eduardo Pérez Pellitero,
Tübingen

# Manifold Learning for Super Resolution

Institut für Informationsverarbeitung
www.tnt.uni-hannover.de

# Manifold Learning for Super Resolution

Von der Fakultät für Elektrotechnik und Informatik

der Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des akademischen Grades

**Doktor-Ingenieur**

genehmigte

**Dissertation**

von

**M. Sc. Eduardo Pérez Pellitero**

geboren am 5. November 1989 in Barcelona.

Betreuer:
Prof. Bodo Rosenhahn, Leibniz Universität Hannover

Gutachter:
Prof. Jörn Ostermann, Leibniz Universität Hannover
Prof. Javier Ruiz-Hidalgo, Universitat Politècnica de Catalunya

Verteidigung der Dissertation: **21.02.2017**

# Fortschritt-Berichte VDI

Reihe 10

Informatik/
Kommunikation

M.Sc. Eduardo Pérez Pellitero,
Tübingen

Nr. 859

# Manifold Learning for Super Resolution

Institut für Informationsverarbeitung
www.tnt.uni-hannover.de

Pérez Pellitero, Eduardo
**Manifold Learning for Super Resolution**
Fortschr.-Ber. VDI Reihe 10 Nr. 859. Düsseldorf: VDI Verlag 2018.
128 Seiten, 47 Bilder, 19 Tabellen.
ISBN 978-3-18-385910-8, ISSN 0178-9627,
€ 48,00/VDI-Mitgliederpreis € 43,20.

**Keywords:** Super Resolution – Manifold Learning – Machine Learning – Computer Vision – Image Processing – Inverse Problems – Upscaling

The development pace of high-resolution displays has been so fast in the recent years that many images acquired with low-end capture devices are already outdated or will be shortly in time. Super Resolution is central to match the resolution of the already existing image content to that of current and future high resolution displays and applications. This dissertation is focused on learning how to upscale images from the statistics of natural images. We build on a sparsity model that uses learned coupled low- and high-resolution dictionaries in order to upscale images, and move towards a more efficient L2 regularization scheme. Instead of using a patch-to-dictionary decomposition, we propose a fully collaborative neighbor embedding approach. We study the positive impact of antipodally invariant metrics for linear regression frameworks, and extend them by also taking into consideration the dihedral group of transforms (i.e. rotations and reflections), as a group of symmetries within the highdimensional manifold. Our experimental section confirms that such approach greatly improves the PSNR performance and also obtains smaller dictionary sizes.

**Bibliographische Information der Deutschen Bibliothek**
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet unter www.dnb.de abrufbar.

**Bibliographic information published by the Deutsche Bibliothek**
(German National Library)
The Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliographie (German National Bibliography); detailed bibliographic data is available via Internet at www.dnb.de.



Als Manuskript gedruckt. Printed in Germany.

ISSN 0178-9627

ISBN 978-3-18-385910-8

## Acknowledgement

Part of the joy of research has a lot to do with the people with whom you share it. I would like to specially thank my doctoral advisor Prof. Dr.-Ing Bodo Rosenhahn for his unconditional support and the always interesting research discussions. Prof. Javier Ruiz has also helped me a lot by always posing challenging questions, being critical but at the same time supportive. I would also like to thank Prof. Dr.-Ing Jörn Ostermann for being part of the defense tribunal.

I keep very good memories from my time at Technicolor, and inevitably some names are linked to them. Special thanks to my research supervisor Dr. Jordi Salvador, who guide me closely during my first steps in research and pushed me always towards the *sunny side* of science. Thanks also to Axel Kochale for being such a helpful person and for the countless times I heard his laughter across the corridor and made me feel a bit happier.

Last but not least, I feel deeply thankful to my parents, who were always there for me, and to my brother Aitor for being such a great brother to look up to.

# Contents

# Abbreviations

| | |
|---|---|
| ANR | Anchored Neighborhood Regression |
| ASRF | Super-Resolution Forest with alternative training |
| CS | Cosine Similarity |
| DCT | Discrete Cosine Transform |
| DLT | Dense Local Training |
| GIBP | Gradient Iterative Back Projection |
| GR | Global Regressor |
| GSM | Gaussian Scale Mixture |
| HHC | Half-Hypersphere Confinement |
| HR | High-Resolution |
| IBP | Iterative Back Projection |
| IFC | Information Fidelity Criterion |
| LASSO | Least Absolute Shrinkage and Selection Operator |
| LR | Low-Resolution |
| MAP | Maximum-a-posteriori |
| MDS | Multidimensional Scaling |
| MF | Magnification Factor |
| ML | Maximum Likelihood |
| MSE | Mean Squared Error |
| NBNN | Naive Bayes Nearest Neighbor |
| NBSRF | Naive Bayes Super-Resolution Forest |
| NN | Nearest Neighbor |
| OMP | Orthogonal Matching Pursuit |
| PCA | Principal Component Analysis |
| PSNR | Peak Signal-to-Noise Ratio |
| PSyCo | Patch Symmetry Collapse |
| SCT | Symmetry-Collapsing Transform |

| | |
|---|---|
| SD | Symmetry Distance |
| SIFT | Scale Invariant Feature Transform |
| SpH | Spherical Hashing |
| SR | Super Resolution |
| SRCNN | Super Resolution using Convolutional Neuronal Networks |
| SSIM | Structural Similarity |
| ST | Symmetry Transform |
| SVD | Singular Value Decomposition |
| VQ | Vector Quantization |

# Abstract

The development pace of high-resolution displays has been so fast in the recent years that many images acquired with low-end capture devices are already outdated or will be shortly in time. Super Resolution is central to match the resolution of the already existing image content to that of current and future high resolution displays and applications. This dissertation is focused on *learning* how to upscale images from the statistics of natural images. We build on a sparsity model that uses learned coupled low- and high-resolution dictionaries in order to upscale images.

Firstly, we study how to adaptively build coupled dictionaries so that their content is semantically related with the input image. We do so by using a Bayesian selection stage which finds the best-fitted texture regions from the training dataset for each input image. The resulting adapted subset of patches is compressed into a coupled dictionary via sparse coding techniques.

We then shift from $\ell_1$ to a more efficient $\ell_2$ regularization, as introduced by Timofte et al. [74]. Instead of using their patch-to-dictionary decomposition, we propose a fully collaborative neighbor embedding approach. In this novel scheme, for each atom in the dictionary we create a densely populated neighborhood from an extensive training set of raw patches (i.e. in the order of hundreds of thousands). This generates more accurate regression functions.

We additionally propose using sublinear search structures such as spherical hashing and trees to speed up the nearest neighbor search involved in regression-based Super Resolution. We study the positive impact of antipodally invariant metrics for linear regression frameworks, and we propose two efficient solutions: (a) the Half Hypersphere Confinement, which enables antipodal invariance within the Euclidean space, and (b) the bimodal tree, whose split functions are designed to be antipodally invariant and which we use in the context of a Bayesian Super Resolution forest.

In our last contribution, we extend antipodal invariance by also taking into consideration the dihedral group of transforms (i.e. rotations and reflections). We study them as a group of symmetries within the high-dimensional manifold. We obtain the respective set of mirror-symmetry axes by means of a frequency analysis, and we use them to collapse the redundant variability, resulting in a reduced manifold span which, in turn, greatly improves quality performance and reduces the dictionary sizes.

# Kurzfassung

In den letzten Jahren ist die Entwicklung von hochauflösenden Displays so rasant verlaufen, dass viele niedrigauflösende Bildaufnahmegeräte entweder schon überholt sind oder es bald sein werden. Super Resolution ist essentiell, um die Auflösung der existierenden Bilder, bzw. Bildinhalte auf die aktuellen und zukünftigen hochauflösenden Displays und Apps zu übertragen. Diese Dissertation beschäftigt sich mit dem lernen des Upscalings von Bildern aus Statistiken realer Bilder. Die Basis hierbei bildet ein *sparsity-Modell*, dass bereits gelernte, gekoppelte hoch- und niedrig auflösende Wörterbücher verwendet, um damit Bilder hochzuskalieren.

Als erstes beschäftigt sich diese Arbeit damit, wie man gekoppelte Wörterbücher adaptativ aufbaut, so, dass ihr Inhalt semantisch mit dem jeweiligen Eingangsbild verknüpft wird. Dies geschieht, indem eine Bayes-Auswahlstufe verwendet wird, welche die am besten passenden Texturbereiche aus dem Trainings-Datensatz für jedes neue Eingangsbild auswählt. Das daraus resultierende angepasste Untermenge von Patches wird anschließend über das Sparse-Coding Techniken in ein gekoppeltes Wörterbuch komprimiert.

Statt einer $\ell_1$ Regularisierung, nutzen wir die effizientere $\ell_2$ Regularisierung, wie von Timofte et al. [74] vorgeschlagen. Anstatt ihrer *patch-to-dictionary*-Zerlegung, wird ein vollstandig kollaborativer Ansatz zur Nachbarschaftseinbettung vorgeschlagen. In diesem neuen Modell schaffen wir für jedes Atom des Wörterbuchs eine dicht besetzte Nachbarschaft die sich aus dem vorherigen umfangreichen Trainingsdatensatz der Roh-Patches ergibt. Dies generiert genauere Regressionsfunktionen.

Zusätzlich wird vorgeschlagen eine sublineare Suchstruktur, z. B. *sperical hashing* und *Bäume* zu nutzen, um die Suche nach dem nächsten Nachbarn zu beschleunigen, die in der regressions-basierten Super-Resolution genutzt wird. Wir analysieren die positiven Auswirkungen der antipodal invarianten Metriken für lineare Regression-frameworks und zwei effiziente Lösungen werden vorgeschlagen: a) das *Half-Hypersphere Confinement*, was die antipode Invarianz innerhalb des Euklidischen Raums ermöglicht sowie b) den *bimodal Baum*, dessen gesplittete Funktionen so angelegt sind, dass sie antipodal invariant sind. Diese werden im dann im Kontext eines „Bayeschen Super Resolution forest“ angewendet.

Als letzten Beitrag wird die antipodale Invarianz ausgeweitet, indem die Diedergruppe der Transformationen (d.h. Rotationen und Reflexionen) ebenso mit einbezogen werden. Diese werden als eine Symmetriegruppe in der hochdimensionalen Mannigfaltigkeit der Patches genauer betrachtet. Die jeweiligen Spiegelsymmetrieachsen werden durch eine Frequenzanalyse bestimmt und anschließend verwendet, um die redundanten Variabilitäten zusammenzufassen. Dies mündet in einer Reduktion des Spannraums der Mannigfaltigkeit, was in einer verbesserten Performanzqualität resultiert und gleichzeitig die Größe des Wörterbuchs verkleinert.