Lie Bodies: A Manifold Representation
of 3D Human Shape
Oren Freifeld1and Michael J. Black2
1Division of Applied Mathematics, Brown University, Providence, RI 02912, USA
2Max Planck Institute for Intelligent Systems, 72076 T¨ ubingen, Germany
freifeld@dam.brown.edu ,black@is.mpg.de
Abstract. Three-dimensional object shape is commonly represented in
terms of deformations of a triangular mesh from an exemplar shape. Ex-
isting models, however, are based on a Euclidean representation of shapedeformat
ions. In contrast, we argue that shape has a manifold structure:
For example, summing the shape deformations for two people does not
necessarily yield a deformation corresponding to a valid human shape,nor does th
e Euclidean di■erence of these two deformations provide a
meaningful measure of shape dissimilarity. Consequently, we de■ne a
novel manifold for shape representation, with emphasis on body shapes,usinganewL
iegroupofdeformations. Thishasseveraladvantages. First
we de■ne triangle deformations exactly, removing non-physical deforma-
tions and redundant degrees of freedom common to previous methods.Second, the Ri
emannian structure of Lie Bodies enables a more mean-
ingful de■nition of body shape similarity by measuring distance between
bodies on the manifold of body shape deformations. Third, the group
structure allows the valid composition of deformations. This is important
for models that factor body shape deformations into multiple causes or
represent shape as a linear combination of basis shapes. Finally,b o d y
shape variation is modeled using statistics on manifolds . Instead of mod-
eling Euclidean shape variation with Principal Component Analysis we
capture shape variation on the manifold using Principal Geodesic Analy-sis. Our
experiments show consistent visual and quantitative advantagesof Lie Bodies over
traditional Euclidean models of shape deformation
and our representation can be easily incorporated into existing methods.
Keywords: Shape deformation, Lie group, Statistics on manifolds.

****

Worldwide Pose Estimation Using 3D Point Clouds■
Yunpeng Li1, Noah Snavely2, Dan Huttenlocher2, and Pascal Fua1
1EPFL
{yunpeng.li,pascal.fua }@epfl.ch
2Cornell University
{snavely,dph }@cs.cornell.edu
Abstract. We address the problem of determining where a photo was taken by
estimating a full 6-DOF-plus-intrincs camera pose with respect to a large geo-
registered 3D point cloud, bringing together research on image localization, lan
d-
mark recognition, and 3D pose estimation. Our method scales to datasets with
hundreds of thousands of images and ten s of millions of 3D points through the
use of two new techniques: a co-occurrence prior for RANSAC and bidirectional
matching of image features with 3D points. We evaluate our method on several
large data sets, and show state-of-the-art results on landmark recognition as we
ll
as the ability to locate camer as to within meters, requiring only seconds per q
uery.

****

Improved Reconstruction of Deforming Surfaces
by Cancelling Ambient Occlusion
Thabo Beeler1,2, Derek Bradley1, Henning Zimmer2, and Markus Gross1,2
1Disney Research Zurich
{derek.bradley,dbeeler }@disneyresearch.com
2ETH Zurich
{hzimmer,grossm }@inf.ethz.ch

Abstract. We present a general technique for improving space-time reconstructionsofdeforming surfaces, whicharecapturedinanvideo-based reconstruction scenar io under uniform illuminati on. Our approach simultaneously improvesboth theacquiredshape as well as thetrackedmotionof the deforming surface. The method is based on factoring out surface shading, computed by a fast approximation to global illumination called ambient occlusion. This allows us to improve the performance of optical■ow tracking that mainly relies on constancy of image features, such as intensity. While cancelling the local shading, we also optimize the surface shape to minimize the residual between the ambient occlusion ofthe 3D geometry and that of the image, yielding more accurate surface details in the reconstruction. Our enhancement is independent of the actual space-time reconstruction algorithm. We experimentally measurethe quantitative improvements produced by our algorithm using a synthetic example of deforming skin, where ground truth shape and motion is available. We further demonstrate our enhancement on a real-worldsequence of human face reconstruction.
1
****

On the Statistical Determination of Optimal Camera Con■gurations in Large Sca le Surveillance Networks

Junbin Liu1, Clinton Fookes1,T i mW a r k2, and Sridha Sridharan1
1Image & Video Research Laboratory, Queensland University of Technology,
2 George Street, Brisbane, QLD 4000, Australia
junbin.liu@gmail.com, {c.fookes,s.sridhara }@qut.edu.au
2CSIRO ICT Centre,
1 Technology Court, Pullenvale, QLD 4069, Australia
tim.wark@csiro.au

Abstract. The selection of optimal camera con■gurations (camera locations, orientations etc.) for multi-camera networks remains an unsolved problem. Previous approaches largely focus on proposing various objective functions to achieve different tasks. Most of them, however, do not generalize well to large scale networks. To tackle this, we introduce a statistical formulation of the optimal selection of camera con■gurations as well as propose a Trans-Dimensional Simulated An-nealing (TDSA) algorithm to effectively solve the problem. We compare our approach with a state-of-the-art method based on Binary Integer Programming (BIP) and show that our approach offers similar performance on small scale problems.However, we also demonstrate the capability of our approach in dealing with large scale problems and show that our approach produces better results than 2 alternative heuristics designed to deal with the scalability issue of BIP.

Keywords: Camera placement, optimization, resersible jump Markov chain Monte Carlo, simulated annealing.
1
****

The Scale of Geometric Texture

Geoffrey Oxholm, Prabin Bariya, and Ko Nishino
Department of Computer Science
Drexel University, Philadelphia, PA 19104, USA
{gao25,pb335,kon }@drexel.edu

Abstract. The most de■ning characteristic of texture is its underlying geometry. Although the appearance of texture is a s dynamic as its illumination and viewing conditions, its geometry remains constant. In this work, we study the fundamental characteristic properties of texture geometry—self similarity and scale variability—and exploit them to perform surface normal estimation, and geometric texture classi■cation. Textures, whether they are regular or stochastic, exhibit some form of repetition in t heir underlying geometry. We use this property to

derive a photometric stereo method uni quely tailored to u tilize the re dundanc
y
in geometric texture. Using basic obser vations about the scal e variability of
tex-
ture geometry, we derive a compact, rotation invariant, scale-space representati
onof geometric texture. To evaluate this representation we introduce an extensiv
e
new texture database that contains multiple distances as well as in-plane and ou
t-
of plane rotations. The high accuracy of the classi■cation results indicate thed
escriptive yet compact nature of our texture representation, and demonstrates
the importance of geometric texture analysis, pointing the way towards improve-
ments in appearance modeling and synthesis.
1
****

E■cient Articulated Trajectory Reconstruction
Using Dynamic Programming and Filters

Jack Valmadre1,3, Yingying Zhu2,3, Sridha Sridharan1, and Simon Lucey1,3
1Queensland University of Technology, Australia
2University of Queensland, Australia
3Commonwealth Scienti■c and Industrial Research Organisation, Australia
{jack.valmadre,yingying.zhu,simon.lucey }@csiro.au

Abstract. This paper considers the problem of reconstructing the mo-
tion of a 3D articulated tree from 2D point correspondences subject
to some temporal prior. Hitherto, smooth motion has been encouraged
using a trajectory basis, yielding a hard combinatorial problem withtime complex
ity growing exponentially in the number of frames. Branch
and bound strategies have previously attempted to curb this complexity
whilst maintaining global optimality. However, they provide no guaran-tee of bei
ng more e■cient than exhaustive search. Inspired by recent
work which reconstructs general trajectories using compact high-pass
■lters, we develop a dynamic programming approach which scales lin-early in the
number of frames, leveraging the intrinsically local nature of
■lter interactions. Extension to a■ne projection enables reconstruction
without estimating cameras.
1
****

Object Co-detection

Sid Yingze Bao, Yu Xiang, and Silvio Savarese
University of Michigan at Ann Arbor, USA
{yingze,yuxiang,silvio}@eecs.umich.edu

Abstract. In this paper we introduce a new problem which we call ob-
ject co-detection . Given a set of images with objects observed from two
or multiple images, the goal of co-detection is to detect the objects, es-tablis
h the identity of individual object instance, as well as estimate the
viewpoint transformation of corresponding object instances. In designing
aco-detector , we follow the intuition that an object has consistent ap-
pearance when observed from the same or di■erent viewpoints. By mod-
eling an object using state-of-the-art part-based representations such as
[1,2], we measure appearance consistency between objects by comparingpart appear
ance and geometry across images. This allows to e■ectively
account for object self-occlusions and viewpoint transformations. Exten-
sive experimental evaluation indicates that our co-detector obtains moreaccurate
 detection results than if objects were to be detected from each
image individually. Moreover, we demonstrate the relevance of our co-
detection scheme to other recognition problems such as single instanceobject rec
ognition, wide-baseline matching, and image query.
1
****

Morphable Displacement Field Based Image

Matching for Face Recognition across Pose
Shaoxin Li1,2, Xin Liu1,2, Xiujuan Chai1, Haihong Zhang3,
Shihong Lao3, and Shiguang Shan1
1Key Lab of Intelligent Information Processing of Chinese Academy of Sciences
(CAS), Institute of Computing Technology, CAS, Beijing, 100190, China
2Graduate University of Chinese Academy of Sciences, Beijing 100049, China
3Omron Social Solutions Co., LTD., Kyoto, Japan
{shaoxin.li,xiujuan.chai,xin.liu,shiguang.shan }@vipl.ict.ac.cn,
lao@ari.ncl.omron.co.jp, angelazhang@ssb.kusatsu.omron.co.jp
Abstract. Fully automatic Face Recognition Across Pose (FRAP) is
one of the most desirable techniques, however, also one of the most chal-
lenging tasks in face recognition ■eld. Matching a pair of face images in
di■erent poses can be converted into matching their pixels correspondingto the s
ame semantic facial point. Following this idea, given two images
GandPin di■erent poses, we propose a novel method, named Mor-
phable Displacement Field (MDF), to match GwithP’s virtual view
underG’s pose. By formulating MDF as a convex combination of a num-
ber of template displacement ■elds generated from a 3D face database,
our model satis■es both global conformity and local consistency. We fur-ther pre
sent an approximate but e■ective solution of the proposed MDF
model, named implicit Morphable Displacement Field (iMDF), which
synthesizes virtual view implicitly via an MDF by minimizing matchingresidual. T
his formulation not only avoids intractable optimization of
the high-dimensional displacement ■eld but also facilitates a constrained
quadratic optimization. The proposed method can work well even whenonly 2 facial
landmarks are labeled, which makes it especially suitable
for fully automatic FRAPsystem. Extensiveevaluations on FERET, PIE
and Multi-PIE databases show considerable improvement over state-of-the-art FRAP
algorithms in both semi-automatic and fully automatic
evaluation protocols.

****

Combining Per-frame and Per-track Cues
for Multi-person Action Recognition
Sameh Khamis, Vlad I. Morariu, and Larry S. Davis
University of Maryland, College Park
{sameh,morariu,lsd }@umiacs.umd.edu
Abstract. We propose a model to combine per-frame and per-track
cues for action recognition. With multiple targets in a scene, our model
simultaneously captures the natural harmony of an individual’s action
in a scene and the ■ow of actions of an individual in a video sequence,inferring
valid tracks in the process. Our motivation is based on the
unlikely discordance of an action in a structured scene, both at the track
level and the frame level (e.g., a person dancing in a crowd of joggers).
While we can utilize sampling approaches for inference in our model, we
instead devise a global inference algorithm by decomposing the problem
and solving the subproblems exactly and e■ciently, recovering a globallyoptimal
joint solution in several cases. Finally, we improve on the state-
of-the-art action recognition results for two publicly available datasets.

****

Joint Image and Word Sense Discrimination
for Image Retrieval
Aurelien Lucchi1,2and Jason Weston1
1Google, New York, USA
2EPFL, Lausanne, Switzerland
Abstract. We study the task of learning to rank images given a text
query,a problem thatis complicated bythe issue of multiple senses. That
is, the senses of interest are typically the visually distinct concepts thata us
er wishes to retrieve. In this paper, we propose to learn a rankingfunction that

optimizes the ranking cost of interest and simultaneously
discovers the disambiguated senses of the query that are optimal for the
supervised task. Note that no supervised information is given about thesenses. E
xperiments performed on web images and the ImageNet dataset
show that using our approach leads to a clear gain in performance.
1
****

Script Data for Attribute-Based
Recognition of Composite Activities
Marcus Rohrbach1, Michaela Regneri2, Mykhaylo Andriluka1,
Sikandar Amin1,3, Manfred Pinkal2, and Bernt Schiele1
1Max Planck Institute for Informatics, Saarbr¨ ucken, Germany
2Department of Computational Linguistics, Saarland University, Germany
3Department of Computer Science, Technische Universit¨ at M¨unchen, Germany
Abstract. State-of-the-art human activity recognition methods build
on discriminative learning which requires a representative training setfor good
performance. This leads to scalability issues for the recognition
of large sets of highly diverse activities. In this paper we leverage the
fact that many human activities are compositional and that the essentialcomponen
ts of the activities can be obtained from textual descriptions
or scripts. To share and transfer knowledge between composite activities
we model them by a common set of attributes corresponding to basicactions and ob
ject participants. This attribute representation allows to
incorporate script data that delivers new variations of a composite ac-
tivity or even to unseen composite activities. In our experiments on 41composite
cooking tasks, we found that script data to successfully cap-
ture the high variability of composite activities. We show improvements
in a supervised case where training data for all composite cooking tasksis avail
able, but we are also able to recognize unseen composites by just
using script data and without any manual video annotation.
1
****

Undoing the Damage of Dataset Bias
Aditya Khosla1, Tinghui Zhou2, Tomasz Malisiewicz1,
Alexei A. Efros2, and Antonio Torralba1
1Massachusetts Institute of Technology
{khosla,tomasz,torralba }@csail.mit.edu
2Carnegie Mellon University
{tinghuiz,efros }@cs.cmu.edu
Abstract. The presence of bias in existing object recognition datasets
is now well-known in the computer vision community. While it remainsin question
whether creating an unbiased dataset is possible given lim-
ited resources, in this work we propose a discriminative framework that
directly exploits dataset bias during training. In particular, our modellearns t
wo sets of weights: (1)bias vectors associated witheach individual
dataset, and (2) visual world weights that are common to all datasets,
which are learned by undoing the associated bias from each dataset. The
visual world weights are expected to be our best possible approximation
to the object model trained on an unbiased dataset, and thus tend to
have good generalization ability. We demonstrate the e■ectiveness of ourmodel by
applying the learned weights to a novel, unseen dataset, and
report superior results for both classi■cation and detection tasks com-
pared to a classical SVM that does not account for the presence of bias.Overall,
we ■nd that it is bene■cial to explicitly account for bias whencombining multip
le datasets.
1
****

Dog Breed Classi■cation Using Part Localization
Jiongxin Liu1, Angjoo Kanazawa2,D a v i dJ a c o b s2, and Peter Belhumeur1
1Columbia University

2University of Maryland
Abstract. We propose a novel approach to ■ne-grained image classi■-
cation in which instances from di■erent classes share common parts but
have wide variation in shape and appearance. We use dog breed identi-
■cation as a test case to show that extracting corresponding parts im-proves cla
ssi■cation performance. This domain is especially challenging
since the appearance of corresponding parts can vary dramatically, e.g.,
the faces of bulldogs and beagles are very di■erent. To ■nd accurate cor-respond
ences, we build exemplar-based geometric and appearance mod-
els of dog breeds and their face parts. Part correspondence allows us to
extract and compare descriptors in like image locations. Our approachalso featur
es a hierarchy of parts (e.g., face and eyes) and breed-speci■c
part localization. We achieve 67% recognition rate on a large real-world
dataset including 133 dog breeds and 8,351 images, and experimentalresults show
that accurate part localization signi■cantly increases clas-si■cation performanc
e compared to state-of-the-art approaches.
1
****

A Dictionary Learning Approach for Classi■cation:
Separating the Particularity and the Commonality■
Shu Kong and Donghui Wang■■
Dept. of Computer Science and Technology, Zhejiang University, Hangzhou, China
{aimerykong,dhwang }@zju.edu.cn
Abstract. Empirically, we ■nd that, despite the class-speci■c features owned by
the objects appearing in the images, the objects from different categories usual
ly
share some common patterns, which do not contribute to the discrimination of
them. Concentrating on this observation and under the general dictionary learnin
g
(DL) framework, we propose a novel met hod to explicitly learn a common pat-
tern pool (the commonality) and class-speci■c dictionaries (the particularity) f
or
classi■cation. We call our method DL-COPAR, which can learn the most com-
pact and most discriminative class-speci■c dictionaries used for classi■cation.
The proposed DL-COPAR is extensively evaluated both on synthetic data and
on benchmark image databases in comparison with existing DL-based classi■-
cation methods. The experimental results demonstrate that DL-COPAR achieves
very promising performances in various applications, such as face recognition,
handwritten digit recognition, scene c lassi■cation and object recognition.
Keywords: Dictionary Learning, Classi■cation, Commonality, Particularity.
1
****

Learning to E■ciently Detect Repeatable
Interest Points in Depth Data
Stefan Holzer1,2,■, Jamie Shotton2, and Pushmeet Kohli2
1Department of Computer Science, CAMP, Technische Universit¨ at M¨unchen (TUM)
holzers@in.tum.de
2Microsoft Research Cambridge
{Jamie.Shotton,pkohli }@microsoft.com
Abstract. Interest point (IP) detection is an important component of
manycomputervision methods.Whilethereare anumberof methodsfor
detecting IPs in RGB images, modalities such as depth images and range
scans have seen relatively little work. In this paper, we approach the IP
detection problem from a machine learning viewpoint and formulate it as
a regression problem. We learn a regression forest (RF) model that, given
animagepatch,tellsusifthereisanIPinthecenterofthepatch.OurRF
basedmethodfor IPdetectionallows aneasy trade-o■betweenspeedand
repeatability by adapting the depth and numberof trees used for approx-
imating the interest point response maps. The data used for training the
RF model is obtained by running state-of-the-art IP detection methods

on the depth images. We show further how the IP response map used for training the RF can be speci■cally designed to increase repeatability by employing 3D models of scenes generated by reconstruction systems such as KinectFusion [1]. Our experiments demonstrate that the use of such data leads to considerably improved IP detection.
1
****
E■ective Use of Frequent Itemset Mining for Image Classi■cation
Basura Fernando1, Elisa Fromont2, and Tinne Tuytelaars1
1KU Leuven, ESAT-PSI, IBBT (Belgium)
2University of Saint-Etienne(France)
Abstract. In this paper we propose a new and e■ective scheme for applying frequent itemset mining to image classi■cation tasks. We refertothenews etofobtainedpatternsas Frequent Local Histograms orFLHs.
During the construction of the FLHs, we pay special attention to keep all the local histogram information during the mining process and to selectthe most relevant reduced set of FLH patterns for classi■cation. The careful choice of the visual primitives and some proposed extensions to exploitothervisualcuessuchas colourorglobal spatial information allowus to build powerful bag-of-FLH -based image representations. We show that these bag-of-FLH s are more discriminative than traditional bag-of-words and yield state-of-the art results on various image classi■cationbenchmarks.
1
****
E■cient Discriminative Projections for Compact Binary Descriptors
Tomasz Trzcinski and Vincent Lepetit
CVLab, EPFL, Lausanne, Switzerland
firstname.lastname@epfl.ch
Abstract. Binary descriptors of image patches are increasingly popular given that they require less storage and enable faster processing.This, however, comes at a price of lower recognition performances. To boost these performances, we project the image patches to a more discriminative subspace, and threshold their coordinates to build our binarydescriptor. However, applying complex projections to thepatches is slow, which negates some of the advantages of binary descriptors. Hence, our key idea is to learn the discriminative projections so that they can bedecomposed into a small number of simple ■lters for which the responses can be computed fast. We show that with as few as 32 bits per descriptor we outperform the state-of-the-art binary descriptors in terms of bothaccuracy and e■ciency.
1
****
Descriptor Learning Using Convex Optimisation
Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman
Visual Geometry Group, University of Oxford
Abstract. The objective of this work is to learn descriptors suitable for the sparse feature detectors used in viewpoint invariant matching. Wemake a number of novel contributions towards this goal: ■rst, it is shown that learning the pooling regions for the descriptor can be formulated as aconvexoptimisation problem selecting the regions using sparsity; second, it is shown that dimensionality reduction can also be formulated as aconvexoptimisation problem, using the nuclear norm to reduce dimensionality. Both of these problems use large margin discriminative learningmethods. The third contribution is a new method of obtaining the positive and negative training data in a weakly supervised manner. And, ■nally, we employ a state-of-the-art stochastic optimizer that is e■cientand well matched to the non-smooth cost functions proposed here. It is

demonstrated that the new learning methods improve over the state of the art in descriptor learning for large scale matching, Brown et al.[2], and large scale object retrieval, Philbin et al.[10].
1
****

Bottom-Up Perceptual Organization of Images into Object Part Hypotheses

Maruthi Narayanan and Benjamin Kimia
Brown University
School of Engineering
Providence, RI 02912
{maruthi
narayanan,benjamin
kimia}@brown.edu
http://vision.lems.brown.edu

Abstract. The demise of “segmentation-then-recognition” strategy led to a paradigm shift toward feature-based discriminative recognition with signi■cant success. However, increased complexity in multi-class datasets reveals that local low-level features may not be su■ciently discriminative, requiring the construction and use of more complex structural features which are necessarily category independent. The paper proposes a bottom-up procedure for generating fragment features which are intended to be object part hypotheses . Suggesting that the demise of segmentation to generate a representation suitable for recognition was due to prematurely committing to a grouping option in the face of ambiguities, the proposed framework considers and tracks multiple alternate grouping options. This approach is made tractable by (i)using amedial fragment representation which allows for the simultaneous use of multiple cues, (ii)a set of transforms to e■ect grouping operations, (iii) acontainment graph representation which avoids duplicate consideration of possibilities, and the estimation of the likelihood of a grouping sequence to retain only plausible groupings. The resulting hypotheses are evaluated intrinsically by measuring their ability to represent objects with a few fragments. They are also evaluated by comparison to algorithms which aim to generate full object segments, with results that match or exceed the state of art, thus demonstrating the suitability of the proposed mid-level representation.
1
****

Match Graph Construction for Large Image Databases

K w a n gI nK i m1, James Tompkin1,2,3,
Martin Theobald1,J a nK a u t z2, and Christian Theobalt1
1Max-Planck-Institut f¨ ur Informatik, Campus E1 4, 66123 Saarbr¨ ucken, Germany
2University College London, Malet Place, WC1E 6BT London, UK
3Intel Visual Computing Institu te, Campus E2 1, 66123 Saarbr¨ ucken, Germany

Abstract. How best to ef■ciently establish correspondence among a large set of images or video frames is an interesting unanswered question. For large databases,the high computational cost of performing pair-wise image matching is a major problem. However, for many applications, images are inherently sparsely connected, and so current techniques try to correctly estimate small potentially matching subsets of databases upon which to perform expensive pair-wise matching. Our contribution is to pose the identi■cation of potential matches as a link prediction problem in an image correspondence graph, and to propose an effective algorithm to solve this problem. Our algorithm facilitates incremental image matching: initially, the match graph is very sparse, but it becomes dense as we alternate between link prediction and veri■cation. We demonstrate the effectivenesof our algorithm by comparing it with several existing alternatives on large-sc

ale
databases. Our resulting match graph is useful for many different applications. As an example, we show the bene■ts of our graph construction method to a labelpropagation application which propagates user-provided sparse object labels to other instances of that object in large image collections.
Keywords: Image matching, graph construction, link prediction.
1
****

Modeling Complex Temporal Composition
of Actionlets for Activity Prediction

Kang Li1,J i eH u2,a n dY u nF u1
1Department of ECE and College of CIS, Northeastern University, Boston, MA, USA
2Department of CSE, State University of New York, Bu■alo, NY, USA
li.ka@husky.neu.edu, y.fu@neu.edu, jhu6@buffalo.edu

Abstract. Early prediction of ongoing activity has been more and more valuable in a large variety of time-critical applications. To build an effective representation for prediction, human activities can be characterized by a complex temporal composition of constituent simple actions.Di■erent from early recognition on short-duration simple activities, wepropose a novel framework for long-duration complex activity prediction
by discovering the causal relationships between constituent actions and
the predictable characteristics of activities. The major contributions ofour work include: (1) we propose a novel activity decomposition method
by monitoring motion velocity which encodes a temporal decomposition
of long activities into a sequence of meaningful action units; (2) Proba-bilistic Su■x Tree (PST) is introduced to represent both large and small
order Markov dependencies between action units; (3) we present a Predictive Accumulative Function (PAF) to depict the predictability of eachkind of activity. The e■ectiveness of the proposed method is evaluated
on two experimental scenarios: activities with middle-level complexity
and activities with high-level complexity. Our method achieves promis-ing results and can predict global activity classes and local action units.
1
****

Learning Human Interaction
by Interactive Phrases

Yu Kong1,3, Yunde Jia1,a n dY u nF u2
1Beijing Laboratory of Intelligent Information Technology
School of Computer Science, Beijing Institute of Technology
Beijing 100081, P.R. China
2Department of ECE and College of CIS, Northeastern University, Boston, MA
3Department of CSE, State University of New York, Bu■alo, NY
{kongyu,jiayunde}@bit.edu.cn,y.fu@neu.edu

Abstract. In this paper, we present a novel approach for human interaction recognition from videos. We introduce high-level descriptions
calledinteractive phrases to express binary semantic motion relationships between interacting people. Interactive phrases naturally exploithuman knowledge to describe interactions and allow us to construct a
more descriptive model for recognizing human interactions. We propose
a novel hierarchical model to encode interactive phrases based on thelatent SVM framework where interactive phrases are treated as latent
variables. The interdependencies between interactive phrases are explicitly captured in the model to deal with motion ambiguity and partialocclusion in interactions. We evaluate our method on a newly collectedBIT-Interaction dataset and UT-Interaction dataset. Promising results
demonstrate the e■ectiveness of the proposed method.
1
****

Learning to Recognize Daily Actions Using Gaze

Alireza Fathi, Yin Li, and James M. Rehg

College of Computing
Georgia Institute of Technology
Abstract. We present a probabilistic generative model for simultaneously recognizing daily actions and predicting gaze locations in videos recorded from an egocentric camera. We focus on activities requiringeye-hand coordination and model the spatio-temporal relationship between the gaze point, the scene objects, and the action label. Our model captures the fact that the distribution of both visual features and ob-ject occurrences in the vicinity of the gaze point is correlated with the verb-object pair describing the action. It explicitly incorporates known properties of gaze behavior from the psychology literature, such as thetemporal delay between ■xation and manipulation events. We present an inference method that can predict the best sequence of gaze locations and the associated action label from an input sequence of images. Wedemonstrate improvements in action recognition rates and gaze predic-tion accuracy relative to state-of-the-art methods, on two new datasets that contain egocentric videos of daily activities and gaze.
1
****
Gait Recognition by Ranking
Ra´ul Mart´■n-F´elez1and Tao Xiang2
1Institute of New Imaging Technologies, Universitat Jaume I, Castell´ o 12071, Spain
2School of EECS, Queen Mary, University of London, London E1 4NS, U.K.
martinr@uji.es, txiang@eecs.qmul.ac.uk
Abstract. The advantage of gait over other biometrics such as face or ■ngerprint is that it can operate from a distance and without subject cooperation. However, this also makes gait subject to changes in variouscovariate conditions including carrying, clothing, surface and view angle. Existing approaches attempt to address these condition changes by feature selection, feature transformation or discriminant subspace learning.However, they su■er from lack of training samples from each subject, canonly cope with changes in a subset of conditions with limited success, and are based on the invalid assumption that the covariate conditions are known a priori. They are thus unable to perform gait recognitionunder a genuine uncooperative setting. We propose a novel approach which casts gait recognition as a bipartite ranking problem and leverages training samples from di■erent classes/people and even from di■er-ent datasets. This makes our approach suitable for recognition under a genuine uncooperative setting and robust against any covariate types, as demonstrated by our extensive experiments.
Keywords: Gait recognition, Learning to rank, Transfer learning.
1
****
Semi-intrinsic Mean Shift on Riemannian Manifolds
Rui Caseiro, Jo˜ ao F. Henriques, Pedro Martins, and Jorge Batista
Institute of Systems and Robotics - University of Coimbra, Portugal
{ruicaseiro,henriques,pedromartins,batista }@isr.uc.pt
Abstract. The original mean shift algorithm [1] on Euclidean spaces (MS) was extended in [2] to operate on general Riemannian manifolds.This extension is extrinsic (Ext-MS) since the mode seeking is performed on thetangent spaces [3], wheretheunderlyingcurvatureisnot fully considered (tangent spaces are only valid in a small neighborhood). In [3]was proposed an intrinsic mean shift designed to operate on two particular Riemannian manifolds (IntGS-MS), i.e. Grassmann and Stiefel manifolds (using manifold-dedicated density kernels). It is then natural to ask whether mean shift could be intrinsically extended to work on alarge class of manifolds. We propose a novel paradigm to intrinsically reformulate the mean shift on general Riemannian manifolds. This is ac-

complished by embedding the Riemannian manifold into a ReproducingKernel Hilbert Space (RKHS) by using a general and mathematically well-founded Riemannian kernel function, i.e. heat kernel [4]. The key issue is that when the data is implicitly mapped to the Hilbert space,the curvature of the manifold is taken into account (i.e. exploits the underlying information of the data). The inherent optimization is then performed on the embedded space. Theoretic analysis and experimentalresults demonstrate the promise and e■ectiveness of this novel paradigm.
1
****

E■cient Nonlocal Regularization for Optical Flow

Philipp Kr¨ ahenb¨uhl and Vladlen Koltun
Stanford University
{philkr,vladlen }@cs.stanford.edu

Abstract. Denseoptical ■ow estimation inimages is achallenging problem because the algorithm must coordinate the estimated motion across large regions in the image, while avoiding inappropriate smoothing over motion boundaries. Recent works have advocated for the use of nonlo-cal regularization to model long-range correlations in the ■ow. However, incorporating nonlocal regularization into an energy optimization framework is challenging due to the large number of pairwise penalty terms.Existing techniques either substitute intermediate ■ltering of the ■ow ■eld for direct optimization of the nonlocal objective, or su■er substantial performance penalties when the range of the regularizer increases. Inthis paper, we describe an optimization algorithm that e■ciently handles a general type of nonlocal regularization objectives for optical ■ow estimation. The computational complexity of the algorithm is indepen-dent of the range of the regularizer. We show that nonlocal regularizationimproves estimation accuracy at longer ranges than previously reported, and is complementary to intermediate ■ltering of the ■ow ■eld. Our algorithm is simple and is compatible with many optical ■ow models.
1
****

Fast Fusion Moves for Multi-model Estimation

Andrew Delong, Olga Veksler, and Yuri Boykov
University of Western Ontario, Canada

Abstract. We develop a fast, effective algorithm for minimizing a well-known objective function for robust multi-model e stimation. Our work introduces a com-
binatorial step belonging to a family of powerful move-making methods likeα-expansion andfusion . We also show that our subproblem can be quickly transformed into a comparatively small instance of minimum-weighted vertex-cover . In practice, these vertex-cover subprobl ems are almost always bipartite and can be solved exactly by specialized network ■ow algorithms. Experiments indicate that our approach achieves the robustness of methods like af■nity propagation, whilst providing the speed of fast greedy heuristics.
1
****

Approximate MRF Inference Using Bounded Treewidth Subgraphs

Alexander Fix1, Joyce Chen1, Endre Boros2,a n dR a m i nZ a b i h1
1Cornell University, Computer Science Department, Ithaca, New York
{afix,yuhsin,rdz }@cs.cornell.edu
2Rutgers University, RUTCOR, New Brunswick, New Jersey
boros@rci.rutgers.edu

Abstract. Graph cut algorithms [9], commonly used in computer vision, solve a ■rst-order MRF over binary variables. T he state of the art for this NP-hard prob-
lem is QPBO [1,2], which ■nds the values for a subset of the variables in the

global minimum. While QPBO is very effective overall there are still many dif■-
cult problems where it can only label a small subset of the variables. We propos
ea new approach that, instead of optimizing the original graphical model, instea
d
optimizes a tractable sub-model , de■ned as an energy function that uses a subse
t
of the pairwise interactions of the original, but for which exact inference can
bedone ef■ciently. Our Bounded Treewidth Subgraph (k-BTS) algorithm greedily
computes a large weight treewidth- ksubgraph of the signed graph, then solves
the energy minimization problem for this subgraph by dynamic programming.The edg
es omitted b y our greedy method provide a p er-instance lower bound. We
demonstrate promising experimental results for binary deconvolution, a challeng-
ing problem used to benchmark QPBO [2]: our algorithm performs an order ofmagnit
ude better than QPBO or its common variants [4], both in terms of energy
and accuracy, and the visual quality of our output is strikingly better as well.
 We
also obtain a signi■cant improvement in energy and accuracy on a stereo bench-ma
rk with 2nd order priors [5], although the improvement in visual quality is
more modest. Our method’s running time is comparable to QPBO.
1
****

Recursive Bilateral Filtering■

Qingxiong Yang
Department of Computer Science,
City University of Hong Kong, Hong Kong, China
http://www.cs.cityu.edu.hk/ ~qiyang/

Abstract. This paper proposes a recursive implementation of the bi-
lateral ■lter. Unlike previous methods, this implementation yields an
bilateral ■lter whose computational complexity is linear in both input
sizeanddimensionality.Theproposedimplementation demonstratesthat
the bilateral ■lter can be as e■cient as the recent edge-preserving ■lter-
ing methods, especially for high-dimensional images. Let the number of
pixels contained in the image be N, and the number of channels be D,
the computational complexity of the proposed implementation will be
O(ND). It is more e■cient than the state-of-the-art bilateral ■ltering
methods that have a computational complexity of O(ND2) [1] (linear
in the image size but polynomial in dimensionality) or O(Nlog(N)D)
[2] (linear in the dimensionality thus faster than [1] for high-dimensional
■ltering). Speci■cally, the proposed implementation takes about 43 ms
to process a one megapixel color image (and about 14 ms to process a
1 megapixel grayscale image) which is about 18 ×faster than [1] and
86×faster than [2]. The experiments were conducted on a MacBook Air
laptop computer with a 1.8 GHz Intel Core i7 CPU and 4 GB memory.
The memory complexity of the proposed implementation is also low: as
few as the image memory will be required (memory for the images be-
fore and after ■ltering is excluded). This paper also derives a new ■lter
namedgradient domain bilateral ■lter from the proposed recursive im-
plementation. Unlike the bilateral ■lter, it performs bilateral ■ltering
on the gradient domain. It can be used for edge-preserving ■ltering but
avoids sharp edges that are observed to cause visible artifacts in some
computer graphics tasks. The proposed implementations were proved to
be e■ective for a number of computer vision and computer graphics ap-
plications, including stylization, tone mapping, detail enhancement and
stereo matching.
1
****

Accelerated Large Scale Optimization
by Concomitant Hashing

Yadong Mu, John Wright, and Shih-Fu Chang
Electrical Engineering Department,

Columbia University, New York, NY 10027
{muyadong,johnwright,sfchang }@ee.columbia.edu
Abstract. Traditional locality-sensitive hashing (LSH) techniques aim to tackle the curse of explosive data scale by guaranteeing that similar samples are projected onto proximal hash buckets. Despite the success of LSH on numerous vision tasks like image retrieval and object matching, however, its potential in large-scale optimization is only realized recently. In this paper we further advance this nascent area. We ■rst identify two common operations known as the computational bottleneck of numerous optimization algorithms in a large-scale setting, i.e., min/max inner product. We propose a hashing scheme for accelerating min/max inner product, which exploits properties of order statistics of statistically correlated random vectors. Compared with other schemes, our algorithm exhibits improved recall at a lower computational cost. The e■ectiveness ande■ciency of theproposedmethodare corroborated bytheoretic analysis and several important applications. Especially, we use the proposed hashing scheme to perform approximate /lscript1regularized least squares with dictionaries with millions of elements, a scale which is beyond the capability of currently known exact solvers. Nonetheless, it is highlighted that the focus of this paper is not on a new hashing scheme for approximate nearest neighbor problem. It exploits a new application for the hashing techniques and proposes a general framework for accelerating a large variety of optimization procedures in computer vision.
1
****

Graph Degree Linkage:
Agglomerative Clustering on a Directed Graph
Wei Zhang1, Xiaogang Wang2,3,D e l iZ h a o1, and Xiaoou Tang1,3
1Department of Information Engineering, The Chinese University of Hong Kong
wzhang009@gmail.com
2Department of Electronic Engineering, The Chinese University of Hong Kong
3Shenzhen Institutes of Adv anced Technology, Chinese A cademy of Sciences, Chin a
Abstract. This paper proposes a simple but effective graph-based agglomerative algorithm, for clustering high-dimensional data. We explore the different roles of
two fundamental concepts in graph theory, indegree and outdegree, in the context of clustering. The average indegree re■ects the density near a sample, and the average outdegree characterizes the local geometry around a sample. Based on such insights, we de■ne the af■nity measure of clusters via the product of average indegree and average outdegree. The product-based af■nity makes our algorithm robust to noise. The algorithm has three main advantages: good performance, easy implementation, and high computational ef■ciency. We test the algorithm on two fundamental computer vision problems: image clustering and object matching. Extensive experiments demonstrate that it outperforms the state-
of-the-arts in both applications.1
1
****

Supervised Earth Mover’s Distance Learning
and Its Computer Vision Applications
Fan Wang and Leonidas J. Guibas
Stanford University, CA, United States
Abstract. The Earth Mover’s Distance (EMD) is an intuitive and natural distance metric for comparing two histograms or probability distri-butions. It provides a distance value as well as a ■ow-network indicating how the probability mass is optimally transported between the bins. In traditional EMD, the ground distance between the bins is pre-de■ned.Instead, we propose to jointly optimize the ground distance matrix and
the EMD ■ow-network based on a partial ordering of histogram dis-

tances in an optimization framework. Our method is further extendedto accept information from general labeled pairs. The trained ground distance better re■ects the cross-bin relationships, hence produces more accurate EMD values and ■ow-networks. Two computer vision applica-tions are used to demonstrate the e■ectiveness of the algorithm: ■rst, we apply the optimized EMD value to face veri■cation, and achieve state-of-the-art performance on the PubFig and the LFW data sets; second, thelearned EMD ■ow-network is used to analyze face attribute changes, obtaining consistent pathsthat demonstrate intuitive transitions on certain facial attributes.

1

****

Global Optimization of Object Pose and Motion from a Single Rolling Shutter Image with Automatic 2D-3D Matching

Ludovic Magerand1,2, Adrien Bartoli2, Omar Ait-Aider1, and Daniel Pizarro3
1Institut Pascal – Université Blaise Pascal – Clermont-Ferrand
2ISIT – Université d'Auvergne – Clermont-Ferrand
3University of Alcala – Alcala de Henares

Abstract. Low cost CMOS cameras can have an acquisition mode called rolling shutter which sequentially exposes the scan-lines. When a single object moves with respect to the camera, this creates image distortions. Assuming 2D-3D correspondences known, previous work showed that theobject pose and kinematics can be estimated from a single rolling shutter image. This was achieved using a suboptimal initialization followed by local iterative optimization.

We propose a polynomial projection model for rolling shutter cameras and a constrained global optimization of its parameters. This isdone by means of a semide■nite programming problem obtained from the generalized problem of moments method. Contrarily to previous work, our optimization does not require an initialization and ensures that theglobal minimum is achieved. This allows us to build automatically robust 2D-3D correspondences using a template to provide an initial set of correspondences.

Experiments show that our method slightly improves previous work on both simulated and real data. This is due to local minima into which previous methods get trapped. We also successfully experimented building 2D-3D correspondences automatically with both simulated and realdata.

Keywords: rolling shutter, motion estimation, matching.

1

****

Online Learning of Linear Predictors for Real-Time Tracking

Stefan Holzer1, Marc Pollefeys2,
Slobodan Ilic1, David Joseph Tan1, and Nassir Navab1
1Department of Computer Science, Technische Universit¨ at M¨unchen (TUM),
Boltzmannstrasse 3, 85748 Garching, Germany
{holzers,slobodan.ilic,tanda,navab }@in.tum.de
2Department of Computer Science, ETH Zurich, CNB G105,
Universitatstrasse 6, CH-8092 Zurich, Switzerland
marc.pollefeys@inf.ethz.ch

Abstract. Although fast and reliable, real-time template tracking using linear predictors requires a long training time. The lack of the ability to learn new templates online prevents their use in applications that require fast learning. This especially holds for applications where the scene is notknown a priori and multiple templates have to be added online. So far, linear predictors had to be either learned o■ine [1] or in an iterative manner by starting with a small sized template and growing it overtime [2]. In this paper, we propose a fast and simple reformulation of the learning procedure that allows learning new linear predictors online.

Keywords: template tracking, template learning, linear predictors.
1
****

Online Learned Discriminative Part-Based
Appearance Models for Multi-human Tracking

Bo Yang and Ram Nevatia
Institute for Robotics and Intelligent Systems,
University of Southern California
Los Angeles, CA 90089, USA
{yangbo,nevatia }@usc.edu

Abstract. We introduce an online learning approach to produce dis-
criminative part-based appearance models (DPAMs) for tracking multi-ple humans i
n real scenes by incorporating association based and cate-
gory free tracking methods. Detection responses are gradually associated
into tracklets in multiple levels to produce ■nal tracks. Unlike most pre-vious
multi-target tracking approaches which do not explicitly consider
occlusions in appearance modeling, we introducea part based model that
explicitly ■nds unoccluded parts by occlusion reasoning in each frame, sothat oc
cluded parts are removed in appearance modeling. Then DPAMs
for each tracklet is online learned to distinguish a tracklet with others
as well as the background, and is further used in a conservative cate-gory free
tracking approach to partially overcome the missed detection
problem as well as to reduce di■culties in tracklet associations under
long gaps. We evaluate our approach on three public data sets, and showsigni■can
t improvements compared with state-of-art methods.
Keywords: multi-humantracking,onlinelearneddiscriminativemodels.
1
****

Exposure Stacks of Live Scenes
with Hand-Held Cameras

Jun Hu1, Orazio Gallo2, and Kari Pulli2
1Department of Computer Science, Duke University
2NVIDIA Research, Santa Clara, CA

Abstract. Many computational photography applications require the
user to take multiple pictures of the same scene with di■erent camera
settings. While this allows to capture more information about the scene
than what is possible with a single image, the approach is limited by therequire
ment that the images be perfectly registered. In a typical scenario
the camera is hand-held and is therefore prone to moving during the
capture of an image burst, while the scene is likely to contain movingobjects. C
ombining such images without careful registration introduces
annoying artifacts in the ■nal image. This paper presents a method to
register exposure stacks in the presence of both camera motion and scenechanges.
Our approach warps and modi■es the content of the images
in the stack to match that of a reference image. Even in the presence
of large, highly non-rigid displacements we show that the images arecorrectly re
gistered to the reference.
1
****

Dual-Force Metric Learning
for Robust Distracter-Resistant Tracker

Zhibin Hong1,X u eM e i2,a n dD a c h e n gT a o1
1Centre for Quantum Computation and Intelligent Systems,
Faculty of Engineering and Information Technology,
University of Technology, Sydney, NSW, Australia
2Center for Automation Research, Electr ical & Computer Engineering Department,
University of Maryland, College Park, MD, USA

Abstract. In this paper, we propose a robust distracter-resistant track-
ing approach by learning a discriminative metric that adaptively learns
the importance of features on-the-■y. The proposed metric is elaboratelydesigned

for the tracking problem by forming a margin objective function which systematically includes distance margin maximization and reconstruction error constraint that acts as a force to push distractersaway from the positive space and into the negative space. Due to the variety of negative samples in the tracking problem, we speci■cally introduce the similarity propagation technique that gives distracters a secondforce from the negative space. Consequently, the discriminative metric obtained helps to preserve the most discriminative information to separate the target from distracters while ensuring the stability of the opti-mal metric. We seamlessly combine it with the popular L1 minimization tracker. Our tracker is therefore not only resistant to distracters, but also inherits the merit of occlusion robustness from the L1 tracker. Quantita-tive comparisons with several state-of-the-art algorithms have been conducted in many challenging video sequences. The results show that our method resists distracters excellently and achieves superior performance.
Keywords: Visual tracking, distracter, distance metric, similarity propagation.
1
****

Shape and Re■ectance from Natural Illumination

Geoffrey Oxholm and Ko Nishino
Department of Computer Science
Drexel University, Philadelphia, PA 19104, USA
{gao25,kon }@drexel.edu

Abstract. We introduce a method to jointly estimate the BRDF and geometry of an object from a single image under known, b ut uncontrolled, natural illumination. We show that this previously unexplored problem becomes tractable when one exploits the orientation clues embedded in the lighting environment. Intuitively, unique regions in the lighting environment act analogously to the point light sources of traditional photometric stereo; they strongly constrain the orientation of the surface patches that re■ect them. The re■ectance, which acts as a bandpass ■lter on the lighting environment, determines the necessary scale ofs uch regions. Accurate re■ectance estimation, however, relies on accurate surface orientation information. Thus, these two factors must be estimated jointly. To do so, we derive a probabilistic formulatio n and introduce priors to address situations where the re■ectance and lighting environment do not suf■ciently constrain the geometry of the object. Through extensive experimentation we show what this space looks like, and offer insights into what problems become solvable inv arious categories of real-world natural illumination environments.
1
****

Frequency Analysis of Transient Light Transport with Applications in Bare Sensor Imaging

Di Wu1,2,5,■, Gordon Wetzstein1, Christopher Barsi1, Thomas Willwacher3, Matthew O’Toole4,N i k h i lN a i k1, Qionghai Dai2, Kyros Kutulakos4, and Ramesh Raskar1
1MIT Media Lab
2Department of Automation, Tsinghua University
3Department of Mathematics, Harvard University
4Department of Computer Science, University of Toronto
5Graduate School at Shenzhen, Tsinghua University

Abstract. Light transport has been analyzed extensively, in both the primal domain and the frequency domain; the latter provides intuition of e■ects introduced by free space propagation and by optical elements, and allows for optimal designs of computational cameras for tailored, e■cient information capture. Here, we relax the common assumption that the speed of light is in■nite and analy ze free space propagation in

the frequency domain considering spatial, temporal, and angular light variation. Using this analysis, we derive analytic expressions for cross-dimensional information transfer and show how this can be exploited for designing a new, time-resolved bare sensor imaging system.
Keywords: Light transport, Fourier analysis, Time of ■ight, Lensless imaging.
1
****

Nonuniform Lattice Regression
for Modeling the Camera Imaging Pipeline
Hai Ting Lin1, Zheng Lu2,
Seon Joo Kim3,a n dM i c h a e lS .B r o w n1
1National University of Singapore
2University of Texas at Austin
3SUNY Korea
Abstract. W ed e s c r i b eam e t h o dt oc o n s t r u c tas p a r s el o o k u pt a b l e
(LUT) that is e■ective in modeling the camera imaging pipeline that maps a RAW camera values to their sRGB output. This work builds on the recent in-camera color processing model proposed by Kim et al. [1] that included a 3D gamut-mapping function. The major drawback in [1] is the high computational cost of the 3D mapping function that uses radial basis functions (RBF) involving several thousand control points. We show how to construct a LUT using a novel nonuniform lattice regression method that adapts the LUT lattice to better ■t the 3D gamut-mapping function. Our method o■ers not only a performance speedup of an order of magnitude faster than RBF, but also a compact mechanism to describe the imaging pipeline.
1
****

Context-Based Automatic Local
Image Enhancement
Sung Ju Hwang1, Ashish Kapoor2, and Sing Bing Kang2
1The University of Texas, Austin,TX, USA
sjhwang@cs.utexas.edu
2Microsoft Research, Redmond, WA, USA
{akapoor,sbkang }@microsoft.com
Abstract. In this paper, we describe a technique to automatically enhance the perceptual quality of an image. Unlike previous techniques, where global statistics of the image are used to determine enhancement operation, our method is local and relies on local scene descriptors andcontext in addition to high-level image statistics. We cast the problem of image enhancement as searching for the best transformation for each pixel in the given image and then discovering the enhanced image usinga formulation based on Gaussian Random Fields. The search is done in a coarse-to-■ne manner, namely by ■nding the best candidate images, followed by pixels. Our experiments indicate that such context-based localenhancement is better than global enhancement schemes. A user study using Mechanical Turk shows that the subjects prefer contextual and local enhancements over the ones provided by existing schemes.
1
****

Segmentation with Non-linear Regional
Constraints via Line-Search Cuts■
Lena Gorelick1, Frank R. Schmidt2,Y u r iB o y k o v1,
Andrew Delong1,a n dA a r o nW a r d1
1University of Western Ontario, Canada
2Universit´ e Paris Est, France
Abstract. This paper is concerned with energy-based image segmentation problems. We introduce a general class of regional functionals

de■ned as an arbitrary non-linear combination of regional unary terms. Such (high-order) functionals are very useful in vision and medical applications and some special cases appear in prior art. For example, our general class of functionals includes but is not restricted to soft constraints on segment volume, its appearance histogram, or shape.
Our overall segmentation energy combines regional functionals with standard length-based regularizers and/or other submodular terms. In general, regional functionals make the corresponding energy minimization NP-hard. We propose a new greedy algorithm based on iterative line search . A parametric max-■ow technique e■ciently explores all solutions along the direction (line) of the steepest descent of the energy.We compute the best “step size”, i.e.the globally optimal solution along the line. This algorithm can make large moves escaping weak local minima, as demonstrated on many real images.
1
****

Hausdor■ Distance Constraint for Multi-surface Segmentation

Frank R. Schmidt1and Yuri Boykov2
1Universit´ e Paris Est, France
2University of Western Ontario, Canada

Abstract. It is well known that multi-surface segmentation can be cast as a multi-labeling problem. Di■erent segments may belong to the same semantic object which may impose various inter-segment constraints [1]. In medical applications, there are a lot of scenarios where upper boundson the Hausdor■ distances between subsequent surfaces are known. We show that incorporating these priors into multi-surface segmentation is potentiallyNP-hard.Tocopewiththisproblemwedevelopasubmodular-supermodular procedure that converges to a locally optimal solution well-approximating the problem. While we cannot guarantee global optimality, only feasible solutions are considered during the optimizationprocess. Empirically, we get useful solutions for many challenging medical applications including MRI and ultrasound images.
1
****

Background Subtraction Using Low Rank and Group Sparsity Constraints

Xinyi Cui1, Junzhou Huang2, Shaoting Zhang1, and Dimitris N. Metaxas1
1CS Dept., Rutgers University
Piscataway, NJ 08854, USA
2CSE Dept., Univ. of Texas at Arlington
Arlington, TX, 76019, USA
{xycui,shaoting,dnm }@cs.rutgers.edu,
jzhuang@uta.edu

Abstract. Background subtraction has been widely investigated in recent years. Most previous work has focused on stationary cameras. Recently, moving cameras have also been studied since videos from mobiledevices have increased signi■cantly. In this paper, we propose a uni■ed and robust framework to e■ectively handle diverse types of videos, e.g., videos from stationary or moving cameras. Our model is inspired by two observations: 1) background motion caused by orthographic cameras lies in a low rank subspace, and 2) pixels belonging to one trajectory tend to group together. Based on these two observations, we introducea new model using both low rank and group sparsity constraints. It is able to robustly decompose a motion trajectory matrix into foreground and background ones. After obtaining foreground and background tra-jectories, the information gathered on them is used to build a statisticalmodel to further label frames at the pixel level. Extensive experiments demonstrate very competitive performance on both synthetic data and real videos.

1
****
F r e eH a n d - D r a w nS k e t c hS e g m e n t a t i o n
Zhenbang Sun1,■, Changhu Wang2, Liqing Zhang1, and Lei Zhang2
1Brain-Like Computing Lab, Shanghai Jiao Tong University, P.R. China
2Microsoft Research Asia
Abstract. In this paper, we study the problem of how to segment a freehand sketch at the object level. By carefully considering the basic principles of human perceptual organization, a real-time solution is presented to automatically segment a user's sketch during his/her drawing. First, a graph-based sketch segmentation algorithm is proposed to segment a cluttered sketch into multiple parts based on the factor of proximity. Then, to improve the ability of detecting semantically meaningful objects, a semantic-based approach is introduced to simulate the past experience in the perceptual system by leveraging a web-scale clipart database. Finally, other important factors learnt from past experience, such assimilarity ,symmetry ,direction ,a n dclosure,a r ea l s ot a k e ni n t o account to make the approach more robust and practical. The proposed sketch segmentation framework has ability to handle complex sketches with overlapped objects. Extensive experimental results show the e■ectiveness of the proposed framework and algorithms.
1
****
Auto-Grouped Sparse Representation for Visual Analysis
Jiashi Feng1, Xiaotong Yuan2, Zilei Wang3,H u a nX u4, and Shuicheng Yan1
1Department of ECE, National University of Singapore
2Department of Statistics, Rutgers University
3Department of Automation, University of Science and Technology of China
4Department of ME, National University of Singapore
Abstract. In this work, we investigate how to automatically uncover the underlying group structure of a feature vector such that each group characterizes certain object-speci■c patterns, e.g.,v i s u a lp a t t e r no rm o - tion trajectories from one object. By mining the group structure, we cane■ectively alleviate the mutual inference of multiple objects and improve the performance in various visual analysis tasks. To this end, we propose a novel auto-grouped sparse representation (ASR) method. ASR groupssemantically correlated feature elements together through optimally fusing their multiple sparse representations. Due to the intractability of primal objective function, we also propose well-behaved convex relaxationand smooth approximation to guarantee obtaining a global optimal solution e■ectively. Finally, we apply ASR in two important visual analysis tasks: multi-label image classi■cation and motion segmentation. Com-prehensive e xperimental evaluations show that ASR is able to achievesuperior performance com pared with the state-of-the-arts on these two tasks.
1
****
A QCQP Approach to Triangulation
Chris Aholt1, Sameer Agarwal2, and Rekha Thomas1
1University of Washington
2Google Inc.
Abstract. Triangulation of a three-dimensional point from n≥2t w o - dimensional images can be formulated as a quadratically constrainedquadratic pro gram. We propose an algorithm to extract candidate solutions to this problem from its semide■nite programming relaxations. We then describe a su■cient condition and a polynomial time test for cer-tifying wh en such a solution is optimal. This test has no false positives.

Experiments indicate that false negatives are rare, and the algorithm has excellent performance in practice. We explain this phenomenon in termsof the geo metry of the triangulation problem.
1
****

Reconstructing the World’s Museums
Jianxiong Xiao1and Yasutaka Furukawa2
1Massachusetts Institute of Technology
2Google Inc.
Abstract. Photorealistic maps are a useful navigational guide for large indoor environments, such as museums and businesses. However, it is impossible to acquire photographs covering a large indoor environment from aerial viewpoints.Thi s paper presents a 3D reconstruction and visualization system to automatically produce clean and well-regularized texture-mapped 3D models for large indoor scenes, from ground-level photographs and 3D laser points. The key component isa new algorithm called “Inverse CSG” for reconstructing a scene in a Constructive Solid Geometry (CSG) representation consisting of volumetric primitives, which imposes powerful regularization constraints to exploit structural regularities. Wealso propose several techniques to adjust the 3D model to make it suitable for rendering the 3D maps from aerial viewpoints. The visualization system enables users to easily browse a large scale indoor environment from a bird’s-eye view,l ocate speci■c room interiors, ■y into a place of interest, view immersive ground-level panorama views, and zoom out again, all with seamless 3D transitions. We demonstrate our system on various museums, including the Metropolitan Mu-seum of Art in New York City – one of the largest art galleries in the world.
1
****

Background Inpainting for Videos with Dynamic Objects and a Free-Moving Camera
Miguel Granados1,K w a n gI nK i m1,J a m e sT o m p k i n1,2,3,
Jan Kautz2, and Christian Theobalt1
1Max-Planck-Institut f¨ ur Informatik, Campus E1 4, 66123 Saarbr¨ ucken, Germany
2University College London, Malet Place, WC1E 6BT London, UK
3Intel Visual Computing Institute, Campus E2 1, 66123 Saarbr¨ ucken, Germany
Abstract. We propose a method for removing marked dynamic objects from videos captured with a free-moving camera, so long as the objects occlude parts of the scene with a static background. Our approach takesas input a video, a mask marking the object to be removed, and a mask marking the dynamic objects to remain in the scene. To inpaint a frame, we align other candidate frames in which parts of the missing region arevisible. Among these candidates, a single source is chosen to ■ll each pixel so that the ■nal arrangement is color-consistent. Intensity di■er-ences between sources are smoothed using gradient domain fusion. Ourframe alignment process assumes that the scene can be approximated using piecewise planar geometry: A set of homographies is estimated for each frame pair, and one each is selected for aligning pixels such that thecolor-discrepancy is minimized and the epipolar constraints are main-tained. We provide experimental validation with several real-world video sequences todemonstrate that, unlikein previouswork, inpainting videosshot with free-moving cameras does not necessarily require estimation of absolute camera positions and per-frame per-pixel depth maps.
Keywords: video processing, video completion, video inpainting, image alignment, background estimation, free-camera, graph-cuts.
1
****

Optimal Templates for Nonrigid Surface Reconstruction
Markus Moll1a n dL u cV a nG o o l1,2
1ESAT-IBBT/PSI - KU Leuven

{markus.moll,luc.vangool }@esat.kuleuven.be
2Computer Vision Laboratory - Swiss Federal Institute of Technology
vangool@vision.ee.ethz.ch
Abstract. This paper addresses the problem of reconstructing a de-
forming surface from point observations in a monocular video sequence.Recentstat
e-of-the-artapproaches dividethesurface intosmaller patches
to simplify the problem. Among these, one very promising approach re-
constructs thepatchesindividually using a quadraticdeformation model.In this pap
er, we demonstrate limitations that a■ect its applicability to
real-world data and propose an approach that overcomes these problems.
In particular, we show how to eliminate the need for manually pickinga template
that is used to model the deformations. We evaluate our al-gorithm on both synth
etic and real-world data sets and show that it
systematically reduces the reconstruction error by a factor of up to ten.
Fig.1.A textured waving ■ag with overlaid vertex mesh of which both the shape an
d
the deformations are recovered
1
****

Learning Domain Knowledge
for Fa¸cade Labelling
Dengxin Dai1,2, Mukta Prasad1, Gerhard Schmitt2, and Luc Van Gool1
1Computer Vision Lab, ETH Z¨ urich
2Chair for Information Architecture, ETH Z¨ urich
Abstract. This paper presents an approach to address the problem of
image fa¸cade labelling. In the architectural literature, domain knowledge
is usually expressed geometrically in the ■nal design, so fa¸ cade labelling
should on the one hand conform to visual evidence, and on the other
hand to the architectural principles – how individual assets (e.g. doors,
windows) interact with each other to form a fa¸ cade as a whole. To this
end,we ■rst propose a recursive splitting method to segment fa¸ cades into
a bunch of tiles for semantic recognition. The segmentation improves
the processing speed, guides visual recognition on suitable scales andrenders th
e extraction of architectural principles easy. Given a set of
segmented training fa¸ cades with their label maps, we then identify a set
ofmeta-featurestocapturebothth isualevidenceandthearchitectural
principles. The features are used to train our fa¸ cade labelling model. In
the test stage, the features are extracted from segmented fa¸ cades and
the inferred label maps. The following three steps are iterated until the
optimal labelling is reached: 1) proposing modi■cations to the current
labelling; 2) extracting new features for the proposed labelling; 3) feeding
the new features to the labelling model to decide whether to accept themodi■cati
ons. In experiments, we evaluated our method on the ECP
fa¸cade dataset and achieved higher precision than the state-of-the-art at
both the pixel level and the structural level.
1
****

Simultaneous Shape and Pose Adaption
of Articulated Models Using Linear
Optimization■
Matthias Straka, Stefan Hauswiesner, Matthias R¨ uther, and Horst Bischof
Institute for Computer Graphics and Vision, Graz University of Technology,
In■eldgasse 16/II, A-8010 Graz, Austria
{straka,hauswiesner,ruether,bischof }@icg.tugraz.at
http://www.icg.tugraz.at/
Abstract. We propose a novel formulation to express the attachment of
a polygonal surface to a skeleton using purely linear terms. This enables
to simultaneously adapt the pose and shape of an articulated model in
an e■cient way. Our work is motivated by the di■culty to constrain a
mesh when adapting it to multi-view silhouette images. However, such

an adaption is essential when capturing the detailed temporal evolution
of skin and clothing of a human actor without markers. While related
work is only able to ensure surface consistency during mesh adaption,
our coupled optimization of the skeleton creates structural stability and
minimizes the sensibilit y to occlusions and out liers ininput images. We
demonstrate the bene■ts of our approach in an extensive evaluation.
The skeleton attachment considerably reduces implausible deformations,
especially when the number of input views is limited.
Keywords: Shape Adaption, Pose Estimation, Mesh Editing, Linear
Optimization.
1
****

Robust Fitting for Multiple View Geometry
Olof Enqvist, Erik Ask, Fredrik Kahl, and Kalle ■Astr¨om
Centre for Mathematical Sciences, Lund University
http://www.maths.lth.se/vision/
Abstract. How hard are geometric vision problems with outliers? We
show that for most ■tting problems, a solution that minimizes the num-ber of out
liers can be found with an algorithm that has polynomial time-
complexity in the number of points (independent of the rate of outliers).
Further, and perhaps more interestingly, other cost functions such as thetruncat
ed L
2-norm can also be handled within the same framework with
the same time complexity.
We apply our framework to triangulation, relative pose problems and
stitching, and give several other examples that ful■ll the required condi-tions.
Based on e■cient polynomial equation solvers, it is experimentally
demonstrated that these problems can be solved reliably, in particular
for low-dimensional models. Comparisons to standard random samplingsolvers are a
lso given.
1
****

Improving Image-Based Localization by Active
Correspondence Search
Torsten Sattler1, Bastian Leibe2, and Leif Kobbelt1
1RWTH Aachen University, Aachen, Germany
2UMIC Research Centre, RWTH Aachen University, Aachen, Germany
Abstract. We propose a powerful pipeline for determining the pose of
a query image relative to a point cloud reconstruction of a large scene
consisting of more than one million 3D points. The key component of oura p p r o
a c hi sa ne ■ c i e n ta n de ■ e c t i v es e a r c hm e t h o dt oe s t a b
l i s hm a t c h e s
between image features and scene points needed for pose estimation.
Our main contribution is a framework for actively searching for addi-tional matc
hes, based on both 2D-to-3D and 3D-to-2D search. A uni■ed
formulation of search in both directions allows us to exploit the distinct
advantages of both strategies, while avoiding their weaknesses. Due toactive sea
rch, the resulting pipeline is able to close the gap in registration
performance observed between e■cient search methods and approaches
that are allowed to run for multiple seconds, without sacri■cing run-time e■cien
cy. Our method achieves the best registration performance
published so far on three standard benchmark datasets, with run-times
comparable or superior to the fastest state-of-the-art methods.
1
****

From Meaningful Contours to Discriminative
Object Shape
Pradeep Yarlagadda and Bj¨ orn Ommer
University of Heidelberg,
Speyererstr. 6, 69115 Heidelberg, Germany

{pradeep.yarlagadda,bjoern.ommer }@iwr.uni-heidelberg.de
Abstract. Shape is a natural, highly prominent characteristic of ob-
jects that human vision utilizes everyday. But despite its expressiveness,
shape poses signi■cant challenges for category-level object detection incluttere
d scenes: Object form is an emergent property that cannot be
perceived locally but becomes only available once the whole object has
been detected and segregated from the background. Thus we address thedetection o
f objects and the assembling of their shape simultaneously.
A dictionary of meaningful contours is obtained by clustering based on
contour co-activation in all training images. We seek a joint, consistentplaceme
nt of all contours in an image, since placing them independently
from another is not reliable due to the emergence of shape. Therefore,
the characteristic object shape is learned by discovering spatially con-sistent
con■gurations of all dictionary contours using maximum margin
multiple instance learning. During recognition, objects are detected and
their shape is explained simultaneously by optimizing a single cost func-tion. W
e demonstrate the bene■t of our approach on standard shape
benchmarks.

****

11. Berg, A.C., Berg, T.L., Malik, J.: Shape matching and object recognition usi
ng
low distortion correspondence. In: CVPR, pp. 26–33 (2005)
12. Julesz, B.: Textons, the elements of texture perception and their interactio
ns. Na-
ture 29(290), 91–97 (1981)
13. Csurka, G., Dance, C.R., Fan, L., Willamowski, J., Bray, C.: Visual categori
zation
with bags of keypoints. In: ECCV, Workshop Stat. Learn. in Comp. Vis. (2004)
14. Gall, J., Lempitsky, V.: Class-speci■c hough forests for object detection. I
n: CVPR
(2009)
15. Maji, S., Malik, J.: Object detection using a max-margin hough transform. In
:
CVPR (2009)
16. Felzenszwalb, P., Girshick, R., McAllester, D., Ramanan, D.: Object detectio
n with
discriminatively trained part-based models. PAMI 32, 1627–1645 (2010)
17. Yarlagadda, P., Monroy, A., Ommer, B.: Voting by Grouping Dependent Parts.
In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part V. LNCS,
vol. 6315, pp. 197–210. Springer, Heidelberg (2010)
18. Gavrila, D.: A bayesian, exemplar-based approach to hierarchical shape match
ing.
PAMI 29 (2007)
19. Felzenszwalb, P., McAllester, D., Ramanan, D.: A discriminatively trained, m
ulti-
scale, deformable part model. In: CVPR (2008)
20. Toshev, A., Taskar, B., Da niilidis, K.: Object detect ion via boundary stru
cture
segmentation. In: CVPR, pp. 950–957 (2010)
21. Zhu, L., Chen, Y., Lin, C., Yuille, A.: Max-margin learning of hierarchical
con■g-
ural deformable templates (hcdt) for e■cient object parsing and pose estimation.
IJCV 93, 1–21 (2011)
22. Fidler, S., Leonardis, A.: Towards scalable representations of object catego
ries:
Learning a hierarchy of parts. In: CVPR (2007)
23. Ahuja, N., Todorovic, S.: Connected segmentation tree: A joint representatio
n of

region layout and hierarchy. In: CVPR (2008)
24. Kokkinos, I., Yuille, A.L.: Hop: Hierarchical object parsing. In: CVPR (2009
)25. Tu, Z., Chen, X., Yuille, A., Zhu, S.: Image parsing: Unifying segmentation
, detec-
tion, and recognition, vol. 2 (2005)
26. Sala, P., Dickinson, S.: Contour Grouping and Abstraction Using Simple Part
Models. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part V.
LNCS, vol. 6315, pp. 603–616. Springer, Heidelberg (2010)
27. Ma, T., Latecki, L.: From partial shape matching through local deformation t
o
robust global shape similarity for object detection. In: CVPR (2011)
28. Srinivasan, P., Zhu, Q., Shi, J.: Many-to-one contour matching for describin
g and
discriminating object shape. In: CVPR (2010)
29. Liu, M., Tuzel, O.: A.Veeraraghavan, Chellappa, R.: Fast directional chamfer
matching. In: CVPR (2010)
30. Andrews, S., Tsochantaridis, I., Hofmann, T.: Support vector machines for
multiple-instance learning. In: NIPS (2003)
31. Narasimhan, M., Bilmes, J.: A submodular-supermodular procedure with applica
-
tions to discriminative structure learning. In: UAI, pp. 401–412 (2005)
32. Riemenschneider, H., Donoser, M., Bischof, H.: Using Partial Edge Contour
Matches for E■cient Object Category Localization. In: Daniilidis, K., Maragos,
P., Paragios, N. (eds.) ECCV 2010, Part V. LNCS, vol. 6315, pp. 29–42. Springer,
Heidelberg (2010)
33. Ferrari,V.,Jurie, F.,Schmid,C.:Fromimagestoshapemodelsforobjectdetection.
IJCV 87, 284–303 (2010)
34. Ommer, B., Malik, J.: Mulit-scale object detection by clustering lines. In:
ICCV
(2009
****
A Particle Filter Framework for Contour Detection
Nicolas Widynski and Max Mignotte
Department of Computer Science and Operations Research (DIRO), University of Mon
treal,
C.P. 6128, succ. Centre-Ville, , Montreal (Quebec), H3C 3J7, Canada
{widynski,mignotte }@iro.umontreal.ca
Abstract. We investigate the contour detection task in complex natural images.
We propose a novel contour detection algorithm which locally tracks small pieces
of edges called edgelets. The combination of the Bayesian modeling and theedgele
ts enables the use of semi-local prior information and image-dependent
likelihoods. We use a mixed of■ine and online learning strategy to detect the mo
st
relevant edgelets. The detection problem is then modeled as a sequential Bayesia
ntracking task, estimated using a particle ■ltering technique. Experiments on th
e
Berkeley Segmentation Datasets show t hat the proposed Particle Filter Contour
Detector method performs well compared to competing state-of-the-art methods.
1
****
TriCoS: A Tri-level Class-Discriminative
Co-segmentation Method for Image Classi■cation
Yuning Chai1,E s aR a h t u2, Victor Lempitsky3,
Luc Van Gool1, and Andrew Zisserman4
1Computer Vision Group, ETH Zurich, Switzerland
2Machine Vision Group, University of Oulu, Finland
3Yandex, Russia
4Visual Geometry Group, University of Oxford, United Kingdom
Abstract. The aim of this paper is to leverage foreground segmentation to improv
e

classi■cation performance on weakly a nnotated datasets – those with no addition
al
annotation other than class labels. We introduce TriCoS, a new co-segmentational
gorithm that looks at all training images jointly and automatically segments out
the most class-discriminative foregrounds for each image. Ultimately, those fore
-
ground segmentations are used to train a classi■cation system.
TriCoS solves the co-segmentation problem by minimizing losses at three dif-
ferent levels: the category level for foreground/background consistency acrossim
ages belonging to the same category, the image level for spatial continuitywithi
n each image, and the dataset level for discrimination between classes.
In an extensive set of experiments, we evaluate the algorithm on three bench-
mark datasets: the UCSD-Caltech Birds-200-2010, the Stanford Dogs, and the
Oxford Flowers 102. With the help of a modern image classi■er, we show su-
perior performance compared to previously published classi■cation methods andoth
er co-segmentation methods.
1
****

Multi-view Discriminant Analysis

Meina Kan1, Shiguang Shan1, Haihong Zhang2,
Shihong Lao2, and Xilin Chen1
1Key Lab of Intelligent Information Processing of Chinese Academy of Sciences
(CAS), Institute of Computing Technology, CAS, Beijing, 100190, China
2Omron Social Solutions Co., LTD., Kyoto, Japan
{meina.kan,shiguang.shan,xilin.chen }@vipl.ict.ac.cn,
angelazhang@ssb.kusatsu.omron.co.jp, lao@ari.ncl.omron.co.jp

Abstract. The same object can be observed at di■erent viewpoints or
evenbydi■erentsensors,thusgenerating multipledistinctevenheteroge-neous samples.
 Nowadays, more and more applications need to recognize
object from distinct views. Some seminal works have been proposed for
object recognition across two views and applied to multiple views insome ine■cie
nt pairwise manner. In this paper, we propose a Multi-view
Discriminant Analysis (MvDA) method, which seeks for a discriminant
common space by jointly learning multiple view-speci■c linear trans-formsforrobu
stobject recognition frommultipleviews,inanon-pairwise
manner. Speci■cally, our MvDA is formulated to jointly solve the multi-
ple linear transforms by optimizing a generalized Rayleigh quotient, i.e.,maximi
zing the between-class variations and minimizing the within-class
variations of the low-dimensional embeddings from both intra-view and
inter-view in the common space. By reformulating this problem as a ra-tio trace
problem, an analytical solution can be achieved by using the
generalized eigenvaluedecomposition. Theproposedmethodisappliedto
three multi-view face recognition problems: face recognition across poses,photo-
sketch face recognition, and Visual (VIS) image vs. Near Infrared
(NIR) image face recognition. Evaluations are conducted respectively on
Multi-PIE, CUFSF andHFB databases. Intensiveexperiments show thatMvDA can achiev
e a more discriminant common space, with up to 13%
improvement compared with the best known results.
Keywords: Multi-view Discriminant Analysis, Multi-view Face Recog-
nition, Common space for Multi-view.
1
****

Multi-scale Patch Based Collaborative
Representation for Face Recognition
with Margin Distribution Optimization

Pengfei Zhu1,L e iZ h a n g1,■,Q i n g h u aH u2, and Simon C.K. Shiu1
1Biometric Research Center, Dept. of Computing,
The Hong Kong Polytechnic University
2School of Computer Science and Technology, Tianjin University
{cspzhu,cslzhang }@comp.polyu.edu.hk

Abstract. Small sample size is one of the most challenging problems in face recognition due to the di■culty of sample collection in many real-world applications. By representing the query sample as a linear combination of training samples from all classes, the so-called collaborative representation based classi■cation (CRC) shows very e■ective face recognition performance with low computational cost. However, the recognition rate of CRC will drop dramatically when the available training samples per subject are very limited. One intuitive solution to this problem is operating CRC on patches and combining the recognition outputs of all patches. Nonetheless, the setting of patch size is a non-trivial task. Considering the fact that patches on di■erent scales can have complementary information for classi■cation, we propose a multi-scalepatchbasedCRCmethod,whiletheensembleofmulti-scaleoutputs is achieved by regularized margin distribution optimization. Our extensive experiments validated that the proposed method outperforms many state-of-the-art patch based face recognition algorithms.
1
****

Object Detection Using
Strongly-Supervised Deformable Part Models
Hossein Azizpour1and Ivan Laptev2
1Computer Vision and Active Perception Laboratory (CVAP), KTH, Sweden
2INRIA, WILLOW, Laboratoire d'Informatique de l'Ecole Normale Superieure
azizpour@kth.se, ivan.laptev@inria.fr
Abstract. Deformable part-based models [1, 2] achieve state-of-the-art performance for object detection, but rely on heuristic initialization dur-ing t raining due to the optimization of non-convex cost function. This paper investigates limitations of such an initialization and extendsearlier methods using additional supervision. We explore strong supervision interms of a nnotated object parts and use it to (i) improve model initialization, (ii) optimize model structure, and (iii) handle partial occlusions. Our method is able to deal with sub-optimal and incomplete annotationsof object parts and is shown to bene■t from semi-supervised learning setups where part-level annotation is provided for a fraction of positive examples only. Experimental results are reported for the detection of sixanimal classes in PASCAL VOC 2007 and 2010 datasets. We demonstrate signi■cant improvements in detection performance compared to the LSVM [1] and the Poselet [3] object detectors.
1
****

E■cient Misalignment-Robust Representation
for Real-Time Face Recognition
Meng Yang, Lei Zhang, and David Zhang
Dept. of Computing, The Hong Kong Polytechnic University, Hong Kong
{csmyang,cslzhang }@comp.polyu.edu.hk
Abstract. Sparse representation techniques for robust face recognition have been widely studied in the past several years. Recently face recognition with simultaneous misalignment, occlusion and other variations has achieved interesting results via robust alignment by sparse representa-tion (RASR). In RASR, the best alignment of a testing sample is sought subject by subject in the database. However, such an exhaustive search strategy can makethetimecomplexity of RASRprohibitiveinlarge-scaleface databases. In this paper, we propose a novel scheme, namely misalignment robust representation (MRR), by representing the misaligned testing sample in the transformed face space spanned by all subjects.The MRR seeks the best alignment via a two-step optimization with a coarse-to-■nesearchstrategy,whichneedsonlytwodeformation-recovery operations. Extensive experiments on representative face databases showthat MRR has almost the same accuracy as RASR in various face recognition and veri■cation tasks but it runs tens to hundreds of times faster

than RASR. The running time of MRR is less than 1 second in thelarge-scale Multi-PIE face database, demonstrating its great potential for real-time face recognition.

1

****

Monocular Object Detection Using 3D Geometric Primitives

Peter Carr1, Yaser Sheikh2, and Iain Matthews1,2
1Disney Research, Pittsburgh
2Carnegie Mellon University

Abstract. Multiview object detection methods achieve robustness in adverse imaging conditions by exploiting projective consistency acrossviews. In this paper, we present an algorithm that achieves performance comparable to multiview methods from a single camera by employing geometric primitives as proxies for the true 3D shape of objects, such aspedestrians or vehicles. Our key insight is that for a calibrated camera,geometric primitives produce predetermined location-speci■c patterns in occupancy maps. We use these to de■ne spatially-varying kernel functions of projected shape. This leads to an analytical formation modelof occupancy maps as the convolution of locations and projected shape kernels. We estimate object locations by deconvolving the occupancy map using an e■cient template similarity scheme. The number of ob-jects and their positions are determined using the mean shift algorithm.

The approach is highly parallel because the occupancy probability of a particular geometric primitive at each ground location is an independentcomputation. The algorithm extends to multiple cameras without requiring signi■cant bandwidth. We demonstrate comparable performance to multiview methods and show robust, realtime object detection on fullresolution HD video in a variety of challenging imaging conditions.

1

****