Resampling Structure from Motion

TianFangandLongQuan
The Hong Kong University of Science and Technology,
Clear Water Bay, Kowloon, Hong Kong, China
{fangtian,quan }@cse.ust.hk

Abstract. This paper proposes a hierarchical framework that resamples 3D reconstructed points to reduce computation cost on time and memory for very large-scale Structure from Motion. The goal is to maintain accuracy and stability similar for different resample rates. We consider this problem in a level-of-detail perspective, from a very large scale global and sparse bundle adjustment to a very detailed and local dense optimization.The dense matching are resampled by exploring the redundancy using local invariant properties, while 3D points are resampled by exploring the redundancy using their covariance and their distribution in both 3Dand image space. Detailed experiments on our resample framework are provided. We also demonstrate the pro posed framework on large-scale examples. The results show that the proposed resample scheme can producea 3D reconstruction with the stability similar to quasi dense methods,while the problem size is as neat as sparse methods.

1

****

Sequential Non-Rigid Structure-from-Motion with the 3D-Implicit Low-Rank Shape Model■

MarcoPaladini1,AdrienBartoli2,andLourdes Agapito1
1Queen Mary University of London, Mile End Road, E1 4NS London, UK
2Clermont Universit´ e, France

Abstract. So far the Non-Rigid Structure-from-Motion problem has been tackled using a batch approach. All the frames are processed at once after the video acquisition takes place. In this paper we propose an incremental approach to the estimation of deformable models. Image frames are processed online in a sequential fashion. The shape is initialised to a rigid model from the first few frames. Subsequently, the problem is formulated as a model based camera tracking problem, where the pose of the camera and the mixing coefficients are updated every frame. New modes are added incrementally when the current model cannot model the current frame well enough. We define a criterion based on image reprojection error to decide whether or not the model must be updated after the arrival of a new frame. The new mode is estimated performing bundle adjustment on a window of frames. To represent the shape, we depart from the traditional explicit low-rank shape model and propose a variant that we call the 3D-implicit low-rank shape model. This alternative model results in a simpler formulation of the motion matrix and provides the ability to represent degenerate deformation modes. We illustrate our approach with experiments on motion capture sequences with ground truth 3D data and with real video sequences.

1

****

Bundle Adjustment in the Large

Sameer Agarwal1,■,N oahSnavely2,StevenM .S eitz3,andRichard Szeliski4
1Google Inc.
2Cornell University
3Google Inc. & University of Washington
4Microsoft Research

Abstract. We present the design and implementation of a new inexact Newton type algorithm for solving large-scale bundle adjustment problems with tens of thousands of images. We explore the use of Conjugate Gradients for calculating the Newton step and its performance as a function of some simple and computationally efficient preconditioners. We show that the common Schur complement trick is not limited to factorization-based methods and that it can be interpreted as a form of preconditioning. Using photos from a street-side dataset and several community photo collections, we generate a variety of bundle adjustment problems and use them to evaluate the performance of six different bundle adjustment algorithms. Our experiments show that truncated Newton methods, when paired with relatively simple preconditioners, offer state of the art performance for large-scale bundle adjustment. The code, test problems and detailed performance data are available athttp://grail.cs.washington.edu/projects/bal .

Keywords: Structure from Motion, Bundle Adjustment, Preconditioned Conjugate Gradients.

****

Sparse Non-linear Least Squares Optimization for Geometric Vision

ManolisI.A. Lo urakis
Institute of Computer Science, Foundation for Research and Technology - Hellas
N. Plastira 100, Vassilika Vouton, Heraklion, Crete, 700 13 Greece
http://www.ics.forth.gr/ ~lourakis/sparseLM/

Abstract. Several estimation problems in vision involve the minimization of cumulative geometric error using non-linear least-squares fit-ting. Typically, this error is characterized by the lack of interdependence among certain subgroups of the parameters to be estimated, which leads to minimization problems possessing a sparse structure. Taking advan-tage of this sparseness during minimization is known to achieve enormous computational savings. Nevertheless, since the underlying sparsity pattern is problem-dependent, its exploitation for a particular estimationproblem requires non-trivial implementation effort, which often discourages its pursuance in practice. Based on recent developments in sparse linear solvers, this paper provides an overview of sparseLM ,a general-purpose software package for sparse non-linear least squares that can exhibit arbitrary sparseness and presents results from its application to important sparse estimation problems in geometric vision.

****

Geometric Image Parsing in Man-Made Environments

OlgaBarinova1,■,V ictorLempitsky2,ElenaTretiak1,andPushmeet Kohli3
1Moscow State University
2University of Oxford
3Microsoft Research Cambridge

Abstract. We present a new parsing framework for the line-based geometric analysis of a single image coming from a man-made environment. This parsing framework models the scene as a composition of geometric primitives spanning different layers from low level (edges) through mid-level (lines and vanishing points) to high level (the zenith and the horizon). The inference in such a model thus jointly and simultaneously estimates a) the grouping of edges into the straight lines, b) the grouping of lines into parallel families, and c) the positioning of the horizon and the zenith in the image. Such a unified treatment means that the uncertainty information propagates between the layers of the model. This is in contrast to most previous approaches to the same problem, which either ignore the middle levels (lines) all together, or use the bottom-up step-by-step pipeline.

For the evaluation, we consider a publicly available York Urban dataset of “Manhattan” scenes, and also introduce a new, harder dataset of 103 urban outdoor images containing many non-Manhattan scenes. The comparative evaluation for the horizon estimation task demonstrate higher accuracy and robustness attained by our method when compared to the current state-of-the-art approaches.

****

Euclidean Structure Recovery from Motion in Perspective Image Sequences via Hankel Rank Minimization

Mustafa Ayazoglu, Mario Sznaier, and Octavia Camps■
Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115, USA

Abstract. In this paper we consider the problem of recovering 3D Euclidean structure from multi-frame point corre spondence data in image sequences under perspective projection. Existing approaches rely either only on geometrical constraints reflecting the rigid nature of the object, or exploit temporal information by recasting the problem into a nonlinear filtering form. In contrast, here we introduce a new constraint th at implicitly exploits the temporal ordering of the frames, leading to a provably correct algorithm to find Euclidean structure (up to a single scaling factor) without the need to alternate between projective depth and motion estimation, estimate the Fundamental matrices or assume a camera motion model. Finally, the proposed approach does not require an accurate calibration of the camera. The accuracy of the algorithm is illustrated using several examples involving both synthetic and real data.

Keywords: Structure from Motion, Perspective Images, Rank Minimization.

1

****

Exploiting Loops in the Graph of Trifocal Tensors for Calibrating a Network of Cameras

J´erˆome Courchay1, Arnak Dalalyan1, Renaud Keriven1, and Peter Sturm2
1IMAGINE, LIGM, Universit´ eParis-Est
2Laboratoire Jean Kuntzmann, INRIA Grenoble Rhˆ one-Alpes

Abstract. A technique for calibrating a network of perspective cameras based on their graph of trifocal tensors is presented. After estimating a set of reliableepipolar geometries, a parameterization of the graph of trifocal tensors is proposed in which each trifocal tensor is encoded by a 4-vector. The strength of this parameterization is that the homographies relating two adjacent trifocal tensors ,as well as the projection matrices depend linearly on the parameters. A method for estimating these parameters in a global way benefiting from loops in the graph is developed. Experiments carried out on several real datasets demonstrate the efficiency of the proposed approach in distributing errors over the whole set of cameras.

1

****

Efficient Structure from Motion by Graph Optimization

MichalHavlena1,AkihikoT orii1,2,andTom´a■sPajdla1
1Center for Machine Perception, Department of Cybernetics, Faculty of Elec. Eng., Czech Technical University in Prague, Technick´ a 2, 166 27 Prague 6, Czech Republic
{havlem1,pajdla }@cmp.felk.cvut.cz
2Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo, Japan
torii@ctrl.titech.ac.jp

Abstract. We present an efficient structure from motion algorithm that can deal with large image collections in a fraction of time and effort of previous approaches while providing comparable quality of the scene and camera reconstruction. First, we employ fast image indexing using largeimage vocabularies to measure visual overlap of images without running actual image matching. Then, we select a small subset from the set of input images by computing its approximate minimal connected dominat-ing set by a fast polynomial algorithm. Finally, we use task prioritization to avoid spending too much time in a few difficult matching problems instead of exploring other easier options. Thus we avoid wasting time onimage pairs with low chance of success and avoid matching of highly redundant images of landmarks. We present results for several challenging sets of thousands of perspective as well as omnidirectional images.

Keywords: Structure from motion, Image set reduction, Task prioritization, Omnidirectional vision.

1

****

Conjugate Gradient Bundle Adjustment

Martin Byr¨odandKalle■Astr¨om■
Centre for Mathematical Sciences, Lund University, Lund, Sweden
{byrod,kalle }@maths.lth.se

Abstract. Bundle adjustment for multi-view reconstruction is traditionally done using the Levenberg-Marquardt algorithm with a direct linear solver, which is computationally very expensive. An alternative to this approach is to apply the conjugate gradients algorithm in the inner loop. This is appealing since the main computational step of the CG algorithm involves only a simple matrix-vector multiplication with the Jacobian. In this work we improve on the latest published approaches to bundle adjustment with conjugate gradients by making full use of the least squares nature of the problem. We employ an easy-to-compute QR factorization based block preconditioner and show how a certain property of the preconditioned system allows us to reduce the work per iteration to roughly half of the standard CG algorithm.

1

****

NF-Features – No-Feature-Features for Representing Non-textured Regions

RalfDragon, M uhammad Shoaib,B odoRosenhahn,andJoernOsterma nn
Institut fuer Informationsverarbeitung
Leibniz Universitaet Hannover
30167 Hannover, Germany
{dragon,shoaib,rosenhahn,ostermann }@tnt.uni-hannover.de

Abstract. In order to achieve a complete image description, we introduce no-feature-features (NF-features) representing object regions where regular interest point detectors do not detect features. As these regionsare usually non-textured, stable re-localization in different images with conventional methods is not possible. Therefore, a technique is presented which re-localizes once-detected NF-features using correspondences of reg-ular features. Furthermore, a distinctive NF descriptor for non-textured regions is derived which has invariance towards affine transformations and changes in illumination. For the matching of NF descriptors, an approachis introduced that is based on local image statistics.

NF-features can be used complementary to all kinds of regular feature detection and description approaches that focus on textured regions, i.e. points, blobs or contours. Using SIFT, MSER, Hessian-Affine or SURF as regular detectors, we demonstrate that our approach is not only suitablefor the description of non-textured areas but that precision and recall of the NF-features is significantly superior to those of regular features. In experiments with high variation of the perspective or image perturbation,at unchanged precision we achieve NF recall rates which are better by more than a factor of two compared to recall rates of regular features.

1

****

Detecting Large Repetitive Structures with Salient Boundaries

ChangchangWu[1],Jan-MichaelFrahm[1],andMarcPollefeys[2]
[1]Department of Computer Science
UNC Chapel Hill, NC, USA
{ccwu,jmf }@cs.unc.edu
[2]Department of Computer Science
ETH Zürich, Switzerland
marc.pollefeys@inf.ethz.ch

Abstract. This paper presents a novel robust and efficient framework to analyze large repetitive structures in urban scenes. A particular contribution of the proposed approach is that it finds the salient boundariesof the repeating elements even when the repetition exists along only one direction. A perspective image is rectified based on vanishing points computed jointly from edges and repeated features detected in the orig-inal image by maximizing its overall symmetry. Then a feature-based method is used to extract hypotheses of repetition and symmetry from the rectified image, and initial repetition regions are obtained from thesupporting features of each repetition interval. To maximize the local symmetry of each element, their boundaries along the repetition direction are determined from the repetition of local symmetry axes. For anyimage patch, we define its repetition quality for each repetition interval conditionally with a suppression of integer multiples of repetition intervals. We determine the boundary along the non-repeating direction byfinding strong decreases of the repetition quality. Experiments demonstrate the robustness and repeatability of our repetition detection.

****

Fast Covariance Computation and Dimensionality Reduction for Sub-window Features in Images

VivekKwatra andMeiHan
Google Research, Mountain View, CA 94043

Abstract. This paper presents algorithms for efficiently computing the covariance matrix for features that form sub-windows in a large multi-dimensional image. For example, several image processing applications, e.g.texture analysis/synthesis, image retrieval, and compression, operate upon patches within an image. These patches are usually projected onto a low-dimensional feature space using dimensionality reduction techniques such as Principal Component Analysis (PCA) and Linear DiscriminantAnalysis (LDA), which in-turn requires computation of the covariance matrix from a set of features. Covariance computation is usually the bottleneck during PCA or LDA ($O(nd^2)$ where $n$is the number of pixels in the image and $d$is the dimensionality of the vector). Our approach reduces the complexity of covariance computation by exploiting the redundancy between feature vectors corresponding to overlapping patches.Specifically, we show that the covariance between two feature components can be reduced to a function of the relative displacement between those components in patch space. One can then employ a lookup tableto store covariance values by relative displacement. By operating in the frequency domain, this lookup table can be computed in $O(n\log n)$ time. We allow the patches to sub-sample the image, which is useful for hierarchical processing and also enables working with filtered responses over these patches, such as local gistfeatures. We also propose a method for fast projection of sub-window patches onto the low-dimensional space.

****

Binary Coherent Edge Descriptors

C.Lawrence Zitnick

Microsoft Research, Redmond, WA

Abstract. Patch descriptors are used for a variety of tasks ranging from finding corresponding points across images, to describing object category parts. In this paper, we propose an image patch descriptor based on edgeposition , orientation and local linear length. Unlike previous works using histograms of gradients, our descriptor does not encode relative gradient magnitudes. Our approach locally normalizes the patch gradients toremove relative gradient information, followed by orientation dependent binning. Finally, the edge histogram is binarized to encode edge locations, orientations and lengths. Two additional extensions are proposedfor fast PCA dimensionality reduction, and a min-hash approach for fast patch retrieval. Our algorithm produces state-of-the-art results on previously published object instance patch data sets, as well as a new patchdata set modeling intra-category appearance variations.

****

Adaptive and Generic Corner Detection Based on the Accelerated Segment Test

Elmar Mair[1,■], Gregory D. Hager[2], Darius Burschka[1],
Michael Suppa[3], and Gerhard Hirzinger[3]

[1] Technische Universität München (TUM), Department of Computer Science,
Boltzmannstr. 3, 85748 Garching bei München, Germany
{elmar.mair,burschka }@cs.tum.edu

[2] Johns Hopkins University (JHU), Department of Computer Science,
3400 N. Charles St., Baltimore, MD 21218-2686, USA
hager@cs.jhu.edu

[3] German Aerospace Center (DLR), Institute of Robotics and Mechatronics,
Münchner Str. 20, 82230 Wessling, Germany
{michael.suppa,gerd.hirzinger }@dlr.de

Abstract. The efficient detection of interesting features is a crucial step for various tasks in Computer Vision. Corners are favored cues due to their two dimensional constraint and fast algorithms to detect them. Recently, a novel corner detection approach, FAST, has been presented which outperforms previous algorithms in both computational performance and repeatability. We will show how the accelerated segment test, which underlies FAST, can be significantly improved by making it more generic while increasing its performance. We do so by finding the optimal decision tree in an extended configuration space, and demonstrating how specialized trees can be combined to yield an adaptive and generic accelerated segment test. The resulting method provides high performance for arbitrary environments and so unlike FAST does not have to be adapted to a specific scene structure. We will also discuss how different test patterns affect the corner response of the accelerated segment test.

Keywords: corner detector, AGAST, adaptive, generic, efficient, AST.

****

Spatially-Sensitive Affine-Invariant Image Descriptors

Alexander M. Bronstein[1,2] and Michael M. Bronstein[1,3]

[1] BBK Technologies ltd.

[2] Dept. of Electrical Engineering, Tel Aviv University

[3] Dept. of Computer Science, Technion – Israel Institute of Technology

Abstract. Invariant image descriptors play an important role in many computer vision and pattern recognition problems such as image searchand retrieval. A dominant paradigm today is that of “bags of features”, a representation of images as distributions of primitive visual elements. The main disadvantage of this approach is the loss of spatial relationsbetween features, which often carry important information about the image. In this paper, we show how to construct spatially-sensitive image descriptors in which both the features and their relation are affine-invariant

. Our construction is based on a vocabulary of pairs of features coupled with a vocabulary of invariant spatial relations between the features. Experimental results show the advantage of our approach in imageretrieval applications.

1

****

Object Classification Using Heterogeneous
Co-occurrence Features
SatoshiItoandSusumu Kubota
Corporate Research & Development Center, Toshiba Corporation, Japan
satoshi13.ito@toshiba.co.jp

Abstract. Co-occurrence features are effective for object classification because observing co-occurrence of two events is far more informative than observing occurrence of each event separately. For example, a colorco-occurrence histogram captures co-occurrence of pairs of colors at a given distance while a color histogram just expresses frequency of each color. As one of such co-occurrence features, CoHOG (co-occurrence his-tograms of oriented gradients) has been proposed and a method using CoHOG with a linear classifier has shown a comparable performance with state-of-the-art pedestrian detection methods. According to recent stud-ies, it has been suggested that combining heterogeneous features such as texture, shape, and color is useful for object classification. Therefore, we introduce three heterogeneous features based on co-occurrencecalled color-CoHOG, CoHED, and CoHD, respectively. Each heterogeneous features are evaluated on the INRIA person dataset and the Oxford 17/102 category flower datasets. The experimental results show thatcolor-CoHOG is effective for the INRIA person dataset and CoHED is effective for the Oxford flower datasets. By combining above heterogeneous features, the proposed method achieves comparable classificationperformance to state-of-the-art methods on the above datasets. The results suggest that the proposed method using heterogeneous features can be used as an off-the-shelf method for various object classification tasks.

1

****

Maximum Margin Distance Learning for
Dynamic Texture Recognition
Bernard Gha nem andNarendraAhuja
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
{bghanem2,ahuja }@vision.ai.uiuc.edu

Abstract. The range space of dynamic textures spans spatiotemporal phenomena that vary along three fundamental dimensions: spatial tex-ture, spatial texture layout, and dynamics. By describing each dimensionwith appropriate spatial or temporal features and by equipping it with a suitable distance measure, elementary distances (one for each dimension) between dynamic texture sequences can be computed. In this paper, weaddress the problem of dynamic texture (DT) recognition by learning linear combinations of these elementary distances. By learning weights to these distances, we shed light on how “salient” (in a discriminative man-ner) each DT dimension is in representing classes of dynamic textures. To do this, we propose an efficient maximum margin distance learning (MMDL) method based on the Pegasos algorithm [1], for both class-independent and class-dependent weight learning. In contrast to popular MMDL methods, which enforce restrictive distance constraints and have a computational complexity that is cubic in the number of training sam-ples, we show that our method, called DL-PEGASOS , can handle more general distance constraints with a computational complexity that can be made linear. When class dependent weights are learned, we showthat, for certain classes of DTs , spatial texture features are dominantly “salient”, while for other classes, this “saliency” lies in their tempo-

ral features. Furthermore, DL-PEGASOS outperforms state-of-the-art recognition methods on the UCLA benchmark DT dataset. By learning class independent weights, we show that this benchmark does not offer much variety along the three DT dimensions, thus, motivating theproposal of a new DT dataset, called DynTex++.

1

****

Image Invariants for Smooth Reflective Surfaces

Aswin C. Sankaranarayanan1, Ashok Veeraraghavan2,
Oncel Tuzel2, and Amit Agrawal2
1Rice University, Houston, TX 77005, USA
2Mitsubishi Electric Research Labs, Cambridge, MA 02139, USA

Abstract. Image invariants are those properties of the images of an object that remain unchanged with changes in camera parameters, illumination etc. In this paper, we derive an image invariant for smooth surfaces with mirror-like reflectance.Since, such surfaces do not have an appearance of their own but rather distort the appearance of the surroundi ng environment, the app licability of geometric invariants is limited. We show that for such smooth mirror-like surfaces, the imagegradients exhibit degeneracy at the surface points that are parabolic. We leverage this result in order to derive a photometric invariant that is associated with parabolic curvature points. Further, we show that these invariant curves can beeffectively extracted from just a few images of the object in uncontrolled, uncalibrated environments without the need for any a priori information about the surface shape. Since these parabolic curves are a geometric property of the surface, they can then be used as features for a variety of machine vision tasks. This is especially powerful, since there are very few vision algorithms that can handle such mirror-like surfaces. We show the potential of the proposed invariant using experiments on two related applications - object recognition and pose estimation for smooth mirror surfaces.

1

****

Visibility Subspaces: Uncalibrated Photometric Stereo with Shadows

Kalyan Sunkavalli, Todd Zickler, and Hanspeter Pfister
Harvard University
33 Oxford St., Cambridge, MA, USA, 02138
{kalyans,zickler,pfister }@seas.harvard.edu

Abstract. Photometric stereo relies on inverting the image formation process, and doing this accurately requires reasoning about the visibilityof light sources with respect to each image point. While simple heuristics for shadow detection suffice in some cases, they are susceptible to error. This paper presents an alternative approach for handling visibility inphotometric stereo, one that is suitable for uncalibrated settings where the light directions are not known. A surface imaged under a finite set of light sources can be divided into regions having uniform visibility, andwhen the surface is Lambertian, these regions generally map to distinct three-dimensional illumination subspaces. We show that by identifying these subspaces, we can locate the regions and their visibilities, and inthe process identify shadows. The result is an automatic method for uncalibrated Lambertian photometric stereo in the presence of shadows, both cast and attached.

1

****

Ring-Light Photometric Stereo

Zhenglong Zhou and Ping Tan
Department of Electrical & Computer Engineering, National University of Singapore

Abstract. We propose a novel algorithm for uncalibrated photometric stereo. While most of previous methods rely on various assumptions on scene properties, we exploit constraints in lighting configurations. We first derive an ambiguous reconstruction by requiring lights to lie on a view centered cone. This reconstruction is upgraded to Euclidean by constraints derived from lights of equal intensity and multiple view geometry. Compared to previous methods, our algorithm deals with more general data and achieves high accuracy. Another advantage of our method is that we can model weak perspective effects of lighting, while previous methods often assume orthographical illumination. We use both synthetic and real data to evaluate our algorithm. We further build a hardware prototype to demonstrate our approach.

1

****

Shape from Second-Bounce of Light Transport

SiyingLiu[1],T ian-TsongNg[1],andYasuyukiMatsush ita[2]
[1]Institute for Infocomm Research Singapore
[2]Microsoft Research Asia

Abstract. This paper describes a method to recover scene geometry from the second-bounce of light transport. We show that form factors (up to a scaling ambiguity) can be derived from the second-bounce component of light transport in a Lambertian case. The form factors carryinformation of the geometric relat ionship between every pair of scene points, i.e., distance between scene points and relative surface orientations. Modelling the scene as polygonal, we develop a method to recoverthe scene geometry up to a scaling ambiguity from the form factors by optimization. Unlike other shape-from-intensity methods, our method simultaneously estimates depth and surface normal; therefore, our methodcan handle discontinuous surfaces as it can avoid surface normal integration. Various simulation and real-world experiments demonstrate the correctness of the proposed theory of shape recovery from light transport.

1

****

A Dual Theory of Inverse and Forward Light Transport

Jiamin Bai[1], Manmohan Chandraker[1],
Tian-Tsong Ng[2], and Ravi Ramamoorthi[1]
[1]University of California, Berkeley
[2]Institute for Infocomm Research, Singapore

Abstract. Inverse light transport seeks to undo global illumination effects, such as interreflections, that pervade images of most scenes. Thispaper presents the theoretical and computational foundations for inverse light transport as a dual of forward rendering. Mathematically, this duality is established through the existence of underlying Neumann se-ries expansions. Physically, we show that each term of our inverse series cancels an interreflection bounce, just as the forward series adds them. While the convergence properties of the forward series are well-known,we show that the oscillatory convergence of the inverse series leads tomore interesting conditions on material reflectance. Conceptually, the inverse problem requires the inversion of a large transport matrix, which is impractical for realistic resolutions. A natural consequence of our the-oretical framework is a suite of fast computational algorithms for light transport inversion – analogous to finite element radiosity, Monte Carlo and wavelet-based methods in forward rendering – that rely at moston matrix-vector multiplications. We demonstrate two practical applications, namely, separation of individual bounces of the light transport and fast projector radiometric compensation to display images free of globalillumina

tion artifacts in real-world environments.

1

****

Lighting Aware Preprocessing for Face
Recognition across Varying Illumination

HuHan1,2,ShiguangShan1,LaiyunQing2, Xilin Chen1,andWenGao1,3
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
3Institute of Digital Media, Peking University, Beijing 100871, China
{hhan,sgshan,lyqing,xlchen,wgao }@jdl.ac.cn

Abstract. Illumination variation is one of intractable yet crucial problems in face recognition and many lighting normalization approacheshave been proposed in the past decades. Nevertheless, most of them preprocess all the face images in the same way thus without considering the specific lighting in each face image. In this paper, we propose a lightingaware preprocessing (LAP) method, which performs adaptive preprocessing for each testing image according to its lighting attribute. Specifically, the lighting attribute of a testing face image is first estimated by usingspherical harmonic model. Then, a von Mises-Fisher (vMF) distribution learnt from a training set is exploited to model the probability that the estimated lighting belongs to normal lighting. Based on this probability,adaptive preprocessing is performed to normalize the lighting variation inthe input image. Extensive experiments on Extended YaleB and Multi-PIE face databases show the effectiveness of our proposed method.

1

****

Detecting Ground Shadows in Outdoor
Consumer Photographs

Jean-Fran¸coisLalonde,A lexeiA .Efros,andSrinivasa G .Narasimhan
School of Computer Science, Carnegie Mellon University
http://graphics.cs.cmu.edu/projects/shadows

Abstract. Detecting shadows from images can significantly improve the performance of several vision tasks such as object detection and tracking. Recent approaches have mainly used illumination invariants whichcan fail severely when the qualities of the images are not very good, as is the case for most consumer-grade photographs, like those on Google or Flickr. We present a practical algorithm to automatically detect shadowscast by objects onto the ground, from a single consumer photograph. Our key hypothesis is that the types of materials constituting the ground in outdoor scenes is relatively limited, most commonly including asphalt,brick, stone, mud, grass, concrete, etc. As a result, the appearances of shadows on the ground are not as widely varying as general shadows and thus, can be learned from a labelled set of images. Our detectorconsists of a three-tier process including (a) training a decision tree classifier on a set of shadow sensitive features computed around each image edge, (b) a CRF-based optimization to group detected shadow edges togenerate coherent shadow contours, and (c) incorporating any existing classifier that is specifically trained to detect grounds in images. Our results demonstrate good detection accuracy (85%) on several challengingimages. Since most objects of interest to vision applications (like pedestrians, vehicles, signs) are attached to the ground, we believe that our detector can find wide applicability.

1

****

The Semi-explicit Shape Model for Multi-object
Detection and Classification■

Simon Pol ak andAmnon S hashua
School of Computer Science and Engineering
The Hebrew University of Jerusalem

Abstract. We propose a model for classification and detection of object classes where the number of classes may be large and where multiple instances of object classes may be present in an image. The algorithm combines a bottom-up, low-level, procedure of a bag-of-words naive Bayes phase for winnowing out unlikely object classes with a high-level procedure for detection and classification. The high-level process is a hybrid of a voting method where votes are filtered using beliefs computed by a class-specific graphical model. In that sense, shape is both explicit (determining the voting pattern) and implicit (each object part votes independently) — hence the term ”semi-explicit shape model”.

1

****

Coupled Gaussian Process Regression for Pose-Invariant Facial Expression Recognition

OgnjenRudovic1,I oannisPatras2,andMajaPantic1,3
1Comp. Dept, Imperial College, London, UK
2Elec. Eng. Dept, Queen Mary University, London, UK
3EEMCS, University of Twente, 7500 AE Enschede, The Netherlands
{o.rudovic,m.pantic }@imperial.ac.uk, i.patras@elec.qmul.ac.uk

Abstract. We present a novel framework for the recognition of facial expressions at arbitrary poses that is based on 2D geometric features. Weaddress the problem by first mapping the 2D locations of landmark points of facial expressions in non-frontal poses to the corresponding locations in the frontal pose. Then, recognition of the expressions is performedby using any state-of-the-art facial expression recognition method (in our case, multi-class SVM). To learn the mappings that achieve pose normalization, we use a novel Gaussian Process Regression (GPR) modelwhich we name Coupled Gaussian Process Regression (CGPR) model.
Instead of learning single GPR model for all target pairs of poses at once, or learning one GPR model per target pair of poses independentlyof other pairs of poses, we propose CGPR model, which also models the couplings between the GPR models learned independently per target pairs of poses. To the best of our knowledge, the proposed method isthe first one satisfying all: (i) being face-shape-model-free, (ii) handling expressive faces in the range from −45°to +45°pan rotation and from −30°to +30°tilt rotation, and (iii) performing accurately for continuous head pose despite the fact that the training was conducted only on a set of discrete poses.

1

****

Bilinear Kernel Reduced Rank Regression for Facial Expression Synthesis

Dong Huang and Fernando De la Torre
Robotics Institute, Carnegie Mellon Unive rsity, Pittsburgh, Pe nnsylvania 15213, USA

Abstract. In the last few years, Facial Expression Synthesis (FES) has been a flourishing area of research driven by a pplications in character animation, computer games, and human computer interaction. This paper proposes a photo-realistic FES method based on Bilinear Kernel Reduced Rank Regression (BKRRR). BKRRR learns a high-dimensional mapping between the appearance of a neutral face and a variety of expressions (e.g. smile, surprise, squint). Thereare two main contributions in this paper: (1) Propose BKRRR for FES. Several algorithms for learning the parameters of BKRRR are evaluated. (2) Propose a new method to preserve subtle person-specific facial characteristics (e.g. wrinkles, pimples). Experimental results on the CMU Multi-PIE database and pictures taken with a regular camera show the effectiveness of our approach.

1

****

Multi-class Classification on Riemannian

Manifolds for Video Surveillance

DiegoT osato[1],M ichelaFarenzena[1],MarcoCristani[1,2],
MauroSpera[1],andVittorio Murino[1,2]
[1]Dipartimento di Informatica, University of Verona, Italy
[2]Istituto Italiano di Tecnologia (IIT), Genova, Italy

Abstract. In video surveillance, classification of visual data can be very hard, due to the scarce resolution and the noise characterizing the sensors' data. In this paper, we propose a novel feature, the ARray of CO-variances (ARCO), and a multi-class classification framework operating on Riemannian manifolds. ARCO is composed by a structure of covariance matrices of image features, able to extract information from data atprohibitive low resolutions. The proposed classification framework consists in instantiating a new multi-class boosting method, working on the manifold Sym
+
dof symmetric positive definite d×d(covariance) matrices. As practical applications, we consider different surveillance tasks, such as head pose classification and pedestrian detection, providing novel state-of-the-art performances on standard datasets.

1

****

Modeling Temporal Structure of Decomposable
Motion Segments for Activity Classification

JuanCarlosNiebles[1,2,3],Chih-WeiChen[1],andLi Fei-Fei[1]
[1]Stanford University, Stanford CA 94305, USA
[2]Princeton University, Princeton NJ 08544, USA
[3]Universidad del Norte, Barranquilla, Colombia

Abstract. Much recent research in human activity recognition has focused on the problem of recognizing simple repetitive (walking, running,waving) and punctual actions (sitting up, opening a door, hugging). However, many interesting human activities are characterized by a complex temporal composition of simple actions. Automatic recognition of suchcomplex actions can benefit from a good understanding of the temporal structures. We present in this paper a framework for modeling motion by exploiting the temporal structure of the human activities. In ourframework, we represent activities as temporal compositions of motion segments. We train a discriminative model that encodes a temporal decomposition of video sequences, and appearance models for each motionsegment. In recognition, a query video is matched to the model according to the learned appearances and motion segment decomposition. Classification is made based on the quality of matching between the motionsegment classifiers and the temporal segments in the query sequence. To validate our approach, we introduce a new dataset of complex Olympic Sports activities. We show that our algorithm performs better than otherstate of the art methods.

Keywords: Activity recognition, discriminative classifiers.

1

****

Cascaded Models for
Articulated Pose Estimation

Benjamin Sapp,A lexanderToshev,andBenTaskar
University of Pennsylvania,
Philadelphia, PA 19104 USA
{bensapp,toshev,taskar }@cis.upenn.edu

Abstract. We address the problem of articulated human pose estimation by learning a coarse-to-fine cascade of pictorial structure models. While the fine-level state-space of poses of individual parts is too largeto permit the use of rich appearance models, most possibilities can be ruled out by efficient structured models at a coarser scale. We propose to learn a sequence of structured models at different pose resolutions,where coar se models filter the pose space for the next level via their max-marginals. The cascade is trained to prune as much as possible while preserving true poses for the final level pictorial structure model. Thefinal level uses much more expensive segmentation, contour and shapefeatures in the model for the remaining filtered set of candidates. We evaluate our framework on the challenging Buffy and PASCAL human pose datasets, improving the state-of-the-art.

****

State Estimation in a Document Image and Its Application in Text Block Identification and Text Line Extraction

HyungIl Koo andNamIkC ho
INMC, Dept. of EECS, Seoul National University
hikoo@ispl.snu.ac.kr ,nicho@snu.ac.kr

Abstract. This paper proposes a new approach to the estimation of document states such as interline spacing and text line orientation, which facilitates a number of tasks in document image processing. The proposedmethod can be applied to spatially varying states as well as invariant ones, so that general cases including images of complex layout, camera-captured images, and handwritten ones can also be handled. Specifically,we find CCs (Connected Components) in a document image and assign a state to each of them. Then the states of CCs are estimated using an energy minimization framework, where the cost function is designed basedon frequency domain analysis and minimized via graph-cuts. Using the estimated states, we also develop a new algorithm that performs text block identification and text line extraction. Roughly speaking, we cansegment an image into text blocks by cutting the distant connections among the CCs (compared to the estimated interline spacing), and we can group the CCs into text lines using a bottom-up grouping along theestimated text line orientation. Experimental results on a variety of document images show that our method is efficient and provides promising results in several document image processing tasks.

Keywords: document image processing, state estimation, graph cuts, text block identification, text line extraction.

****

Discriminative Learning with Latent Variables for Cluttered Indoor Scene Understanding

HuayanWang1,StephenGould2,andDaphneKoller1
1Computer Science Department, Stanford University, CA, USA
2Electrical Engineering Department, Stanford Univeristy, CA, USA

Abstract. We address the problem of understanding an indoor scene from a single image in terms of recovering the layouts of the faces (floor, ceiling, walls) and furniture. A major challenge of this task arises from the fact that most indoor scenes are cluttered by furniture and decora-tions, whose appearances vary drastically across scenes, and can hardly be modeled (or even hand-labeled) consistently. In this paper we tackle this problem by introducing latent variables to account for clutters, sothat the observed image is jointly explained by the face and clutter layouts. Model parameters are learned in the maximum margin formulation, which is constrained by extra prior energy terms that define the role ofthe latent variables. Our approach enables taking into account and in-ferring indoor clutter layouts without hand-labeling of the clutters in the training set. Yet it outperforms the state-of-the-art method of Hedau et al. [4] that requires clutter labels.

****

Simultaneous Segmentation and Figure/Ground Organization Using Angular Embedding

MichaelMaire
California Institute of Technology - Pasadena, CA, 91125
mmaire@caltech.edu

Abstract. Image segmentation and figure/ground organization are fundamental steps in visual perception. This paper introduces an algorithmthat couples these tasks together in a single grouping framework driven by low-level image cues. By encoding both affinity and ordering preferences in a common representation and solving an Angular Embeddingproblem, we allow segmentation cues to influence figure/ground assignment and figure/ground cues to influence segmentation. Results are comparable to state-of-the-art automatic image segmentation systems, whileadditionally providing a global figure/ground ordering on regions.

1

****

Cosegmentation Revisited:
Models and Optimization

SaraVicente1,V ladimirKolmogorov1,andCarstenRother2
1University College London
2Microsoft Research Cambridge

Abstract. The problem of cosegmentation consists of segmenting the same object (or objects of the same class) in two or more distinct im-ages. Recently a number of different models have been proposed for this problem. However, no comparison of such models and corresponding optimization techniques has been done so far. We analyze three existingmodels: the L1 norm model of Rother et al. [1], the L2 norm model of Mukherjee et al. [2] and the “reward” model of Hochbaum and Singh [3]. We also study a new model, which is a straightforward extension of theBoykov-Jolly model for single image segmentation [4].

In terms of optimization, we use a Dual Decomposition (DD) technique in addition to optimization methods in [1,2]. Experiments show a significant improvement of DD over published methods. Our main conclusion, however, is that the new model is the best overall because it: (i)has fewest parameters; (ii) is most robust in practice, and (iii) can be optimized well with an efficient EM-style procedure.

1

****

Optimal Contour Closure by
Superpixel Grouping

Alex Levinshtein1, Cristian Sminchisescu2, and Sven Dickinson1
1University of Toronto
{babalex,sven }@cs.toronto.edu
2University of Bonn
cristian.sminchisescu@ins.uni-bonn.de

Abstract. Detecting contour closure, i.e., finding a cycle of disconnected contour fragments that separates an object from its background,is an important problem in perceptual grouping. Searching the entire space of possible groupings is intractable, and previous approaches have adopted powerful perceptual grouping heuristics, such as proximity andco-curvilinearity, to manage the search. We introduce a new formulation of the problem, by transforming the problem of finding cycles of contour fragments to finding subsets of superpixels whose collective boundaryhas strong edge support in the image. Our cost function, a ratio of a novel learned boundary gap measure to area, promotes spatially coherent sets of superpixels. Moreover, its properties support a global optimiza-tion procedure using parametric maxflow. We evaluate our framework by comparing it to two leading contour closure approaches, and find that it yields improved performance.

1

****

Fast and Exact Primal-Dual Iterations for Variational Problems in Computer Vision

Jan Lellmann, Dirk Breitenreicher, and Christoph Schn¨ orr
Image and Pattern Analysis Group & HCI
Dept. of Mathematics and Computer Science, University of Heidelberg
{lellmann,breitenreicher,schnoerr }@math.uni-heidelberg.de

Abstract. The saddle point framework provides a convenient way to formulate many convex variational problems that occur in computer vi-sion. The framework unifies a broad range of data and regularization terms, and is particularly suited for nonsmooth problems such as Total Variation-based approaches to image labeling. However, for manyinteresting problems the constraint sets involved are difficult to handle numerically. State-of-the-art methods rely on using nested iterative projections, which induces both theoretical and practical convergence is-sues. We present a dual multiple-constraint Douglas-Rachford splitting approach that is globally convergent, avoids inner iterative loops, enforces the constraints exactly, and requires only basic operations thatcan be easily parallelized. The method outperforms existing methods by af a c t o ro f4 –20 while considerably increasing the numerical robustness.

1

****

An Experimental Study of Color-Based Segmentation Algorithms Based on the Mean-Shift Concept

K. Bitsakos, C. Fermüller, and Y. Aloimonos
Center for Automation Research,
University of Maryland, College Park, USA
kbits@cs.umd.edu, {fer,yiannis}@cfar.umd.edu

Abstract. We point out a difference between the original mean-shift formulation of Fukunaga and Hostetler and the common variant in the computer vision community, namely whether the pairwise comparison is performed with the original or with the filtered image of the previousiteration. This leads to a new hybrid algorithm, called Color Mean Shift, that roughly speaking, treats color as Fukunaga’s algorithm and spatial coordinates as Comaniciu’s algorithm. We perform experiments toevaluate how different kernel functions and color spaces affect the final filtering and segmentation results, and the computational speed, using the Berkeley and Weizmann segmentation databases. We conclude thatthe new method gives better results than existing mean shift ones on four standard comparison measures (/revsimilar15%,22%improvement on RAND and BDE measures respectively for color images), with slightly higher run-ning times (/revsimilar10%). Overall, the new method produces segmentations comparable in quality to the ones obtained with current state of the art segmentation algorithms.

Keywords: image segmentation, image filtering, mean-shift.

1

****

Towards More Efficient and Effective LP-Based Algorithms for MRF Optimization

Nikos Komodakis
University of Crete
Computer Science Department
komod@csd.uoc.gr

Abstract. This paper proposes a framework that provides significant speed-ups and also improves the effectiveness of general message passingalgorithms based on dual LP relaxations. It is applicable to both pairwise and higher order MRFs, as well as to any type of dual relaxation. It relies on combining two ideas. The first one is inspired by algebraicmultigrid approaches for linear systems, while the second one employsa novel decimation strategy that carefully fixes the labels for a growing subset of nodes during the course of a dual LP-based algorithm. Experimental results on a wide variety of vision problems demonstrate the great effectiveness of this framework.

1

****

Energy Minimization under Constraints on Label Counts

Yongsub Lim[1], Kyomin Jung[1],■, and Pushmeet Kohli[2]
[1]Korea Advanced Institute of Science and Technology, Daejeon, Korea
yongsub@kaist.ac.kr ,kyomin@kaist.edu
[2]Microsoft Research, Cambridge, United Kingdom
pkohli@microsoft.com

Abstract. Many computer vision problems such as object segmentation or reconstruction can be formulated in terms of labeling a set of pixels or voxels. In certain scenarios, we may know the number of pixels or voxels which can be assigned to a particular label. For instance, in the reconstruction problem, we may know size of the object to be reconstructed. Such label count constraints are extremely powerful and have recently been shown to result in good solutions for many vision problems.

Traditional energy minimization algorithms used in vision cannot handle label count constraints. This paper proposes a novel algorithm for minimizing energy functions under constraints on the number of variables which can be assigned to a particular label. Our algorithm is deterministic in nature and outputs ε-approximate solutions for all possible counts of labels. We also develop a variant of the above algorithm which is much faster, produces solutions under almost all label count constraints, and can be applied to all submodular quadratic pseudo-boolean functions. We evaluate the algorithm on the two-label (foreground/background) image segmentation problem and compare its performance with the state-of-the-art parametric maximum flow and max-sum diffusion based algorithms. Experimental results show that our method is practical and is able to generate impressive segmentation results in reasonable time.

1

****

A Fast Dual Method for HIK SVM Learning

Jianxin Wu■
School of Computer Engineering, Nanyang Technological University
jxwu@ntu.edu.sg

Abstract. Histograms are used in almost every aspect of computer vision, from visual descriptors to image representations. Histogram Intersection Kernel (HIK) and SVM classifiers are shown to be very effective in dealing with histograms. This paper presents three contributions concerning HIK SVM classification. First, instead of limited to integer histograms, we present a proof that HIK is a positive definite kernel for non-negative real-valued feature vectors. This proof reveals some interesting properties of the kernel. Second, we propose ICD, a deterministic and highly scalable dual space HIK SVM solver. ICD is faster than and has similar accuracies with general purpose SVM solvers and two recently proposed stochastic fast HIK SVM training methods. Third, we empirically show that ICD is not sensitive to the Cparameter in SVM. ICD achieves high accuracies using its default parameters in many datasets. This is a very attractive property because many vision problems are too large to choose SVM parameters using cross-validation.

1

****

Weakly-Paired Maximum Covariance Analysis for Multimodal Dimensionality Reduction and Transfer Learning

Christoph H. Lampert[1]and Oliver Kr¨ omer[2]
[1]Institute of Science and Technology Austria, Klosterneuburg, Austria
[2]Max Planck Institute for Biological Cybernetics, T¨ ubingen, Germany

Abstract. We study the problem of multimodal dimensionality reduction assuming that data samples can be missing at training time, and not all data modalities may be present at application time. Maximum covariance analysis , as a generalization of PCA, has many desirable properties, but its application to practical problems is limited by its need for perfectly paired data. We overcome this limitation by a latent variableapproach that allows working with weakly paired data and is still able to efficiently process large datasets usi ng standard numerical routines. The resulting weakly paired maximum covariance analysis often finds better representations than alternative methods, as we show in two exemplarytasks: texture discrimination and transfer learning.

1

****

Optimizing Complex Loss Functions in Structured Prediction

Mani Ranjbar, Greg Mori, and Yang Wang
School of Computing Science
Simon Fraser University, Canada

Abstract. In this paper we develop an algorithm for structured prediction that optimizes against complex performance measures, those which are a function of false positive and false negative counts. The approach can be directly applied to performance measures such as Fβscore (natural language processing), intersection over union (image segmentation),Precision/Recall at k (search engines) and ROC area (binary classifiers).
We attack this optimization problem by approximating the loss function with a piecewise linear function and relaxing the obtained QP problemto a LP which we solve with an off-the-shelf LP solver. We present experiments on object class-specific segmentation and show significant improvement over baseline approaches that either use simple loss functionsor simple compatibility functions on VOC 2009.

1

****

A Novel Parameter Estimation Algorithm for the Multivariate t-Distribution and Its Application to Computer Vision

Chad Aeschliman, Johnny Park, and Avinash C. Kak
Purdue University
http://rvl.ecn.purdue.edu

Abstract. We present a novel algorithm for approximating the parameters of a multivariate t-distribution. At the expense of a slightly decreased accuracy in the estimates, the proposed algorithm is significantly faster and easier to implement compared to the maximum likelihood estimates computed using the expectation-maximization algorithm. The formulation of the proposed algorithm also provides theoretical guidance for solving problems that are intractable with the maximum likelihood equations. In particular, we show how the proposed algorithm can be modified to give an incremental solution for fast online parameter estimation. Finally, we validate the effectiveness of the proposed algorithm by using the approximated t-distribution as a drop in replacement for the conventional Gaussian distribution in two computer vision applications: object recognition and tracking. In both cases the t-distribution gives better performance with no increase in computation.

1

****

LACBoost and FisherBoost: Optimally Building Cascade Classifiers

Chunhua Shen[1,2], Peng Wang[3,■], and Hanxi Li[2,1]
[1]NICTA■■, Canberra Research Laboratory, ACT 2601, Australia
[2]Australian National University, ACT 0200, Australia
[3]Beihang University, Beijing 100191, China

Abstract. Object detection is one of the key tasks in computer vision. The cascade framework of Viola and Jones has become the de facto standard. A classifier in each node of the cascade is required to achieve extremely high detection rates, inst ead of low overall classification error. Although there are a few reported methods addressing this requirement in the context of object detection, there is no a principled feature selection method that explicitly takes into account this asymmetric node learning objective. We provide such a boosting algorithm in this work. It is inspired by the linear asymmetric classifier (LAC) of [1] in that our boosting algorithm optimizes a similar cost function. The new totally-corrective boosting algorithm is implemented by the column generation technique in convex optimization. Experimental results on face detection suggest that our proposed boosting algorithms can improve the state-of-the-art methods in detection performance.

****

A Shrinkage Learning Approach for Single Image Super-Resolution with Overcomplete Representations

Amir Adler[1], Yacov Hel-Or[2], and Michael Elad[1]
[1]Computer Science Department, The Technion, Haifa, Israel
[2]Efi Arazi School of Computer Science,
The Interdisciplinary Center, Herzelia, Israel

Abstract. We present a novel approach for online shrinkage functions learning in single image super-resolution. The proposed approach leverages the classical Wavelet Shrinkage denoising technique where a set of scalar shrinkage functions is applied to the wavelet coefficients of a noisy image. In the proposed approach, a unique set of learned shrinkage functions is applied to the overcomplete representation coefficients of the interpolated input image. The super-resolution image is reconstructed from the post-shrinkage coefficients. During the learning stage, the low-resolution input image is treated as a reference high-resolution image and a super-resolution reconstruction process is applied to a scaled-down version of it. The shapes of all shrinkage functions are jointly learned by solving a Least Squares optimization problem that minimizes the sum of squared errors between the reference image and its super-resolution approximation. Computer simulations demonstrate superior performance compared to state-of-the-art results.

****

Object of Interest Detection by Saliency Learning

Pattaraporn Khuwuthyakorn[1,3], Antonio Robles-Kelly[1,2], and Jun Zhou[1,2]
[1]RSISE, Australian National University, Canberra, ACT 0200, Australia
[2]National ICT Australia (NICTA■), Canberra, ACT 2601, Australia
[3]Cooperative Research Centre for National Plant Biosecurity■■,
Canberra, ACT, 2617, Australia

Abstract. In this paper, we present a method for object of interest detection. This method is statistical in nature and hinges in a model which combines salient features using a mixture of linear support vector machines. It exploits a divide-and-conquer strategy by partitioning the feature space into sub-regions of linearly separable data-points. This yields a structured learning approach where we learn a linear support vector machine for each region, the mixture weights, and the combination parameters for each of the salient features at hand. Thus, the method learns the combination of salient features such that a mixture of classifiers can be used to recover objects of interest in the image. We illustrate the utility of the method by applying our algorithm to the MSRA Salient Object Database.

1

****

Boundary Detection Using F-Measure-, Filter- and Feature- (F3) Boost

Iasonas Kokkinos
Department of Applied Mathematics, Ecole Centrale Paris
INRIA-Saclay, GALEN Group

Abstract. In this work we propose a boosting-based approach to boundary detection that advances the current state-of-the-art. To achieve this we introduce the following novel ideas: (a) we use a training criterion that approximates the F-measure of the classifier, instead of the exponential loss that is commonly used in boosting. We optimize this criterion using Anyboost. (b) We deal with theambiguous information about orientation of the boundary in the annotation by treating it as a hidden variable, and train our classifier using Multiple-Instance Learning. (c) We adapt the F ilterboost approach of [1] to leverage information from the whole training set to train our classifier, instead of using a fixed subset of points. (d) We extract discriminative features from appearance descriptors that are computed densely over the image. We demonstrate the performance of ourapproach on the Berkeley Segmentation Benchmark.

1

****

Unsupervised Learning of Functional Categories in Video Scenes■

Matthew W. Turek, Anthony Hoogs, and Roderic Collins
Kitware, Inc., Clifton Park, N.Y. U.S.A.
{matt.turek,anthony.hoogs,roddy.collins}@kitware.com
http://www.kitware.com

Abstract. Existing methods for video scene analysis are primarily concerned with learning motion patterns or models for anomaly detection. We present a novel form of video scene analysis where scene element categories such as roads, parking areas, sidewalks and entrances, can be segmented and categorized based on the behaviors of moving objects in and around them. We view the problem from the perspective of categorical object recognition, and present an approach for unsupervised learning offunctional scene element categories. Our approach identifies functional regions with similar behaviors in the same scene and/or across scenes, by clustering histograms based on a trajectory-level, behavioral codebook. Experiments are conducted on two outdoor webcam video scenes with low frame rates and poor quality. Unsupervised classification results are presented for each scene independently, and also jointly where models learned on one scene are applied to the other.

Keywords: functional modeling, unsupervised learning, video analysis.

1

****

Automatic Learning of Background Semantics in Generic Surveilled Scenes

Carles Fern´ andez, Jordi Gonz` alez, and Xavier Roca
Dept. Ci` encies de la Computaci´ o & Computer Vision Center,
Edifici O, Campus UAB, 08193 Bellaterra, Barcelona, Spain

{carles.fernandez,poal,xavier.roca }@cvc.uab.es

Abstract. Advanced surveillance systems for behavior recognition in outdoor traffic scenes depend strongly on the particular configuration of the scenario. Scene-independent trajectory analysis techniques statistically infer semantics in locations where motion occurs, and suchinferences are typically limited to abnormality. Thus, it is interestingto design contributions that automatically categorize more specific semantic regions. State-of-the-art approaches for unsupervised scene labeling exploit trajectory data to segment areas like sources, sinks, orwaiting zones. Our method, in addition, incorporates scene-independent knowledge to assign more meaningful labels like crosswalks, sidewalks, or parking spaces. First, a spatiotemporal scene model is obtained fromtrajectory analysis. Subsequently, a so-called GI-MRF inference process reinforces spatial coherence, and incorporates taxonomy-guided smoothness constraints. Our method achieves automatic and effective labelingof conceptual regions in urban scenarios, and is robust to tracking errors. Experimental validation on 5 surve illance databases has been conducted to assess the generality and accuracy of the segmentations. The resultingscene models are used for model-based behavior analysis.

1

****

Why Did the Person Cross the Road (There)?
Scene Understanding Using Probabilistic Logic
Models and Common Sense Reasoning

Aniruddha Kembhavi, Tom Yeh, and Larry S. Davis
University of Maryland, College Park
anikem@umd.edu, tomyeh@umiacs.umd.edu, lsd@cs.umd.edu

Abstract. We develop a video understanding system for scene elements, such as bus stops, crosswalks, and intersections, that are characterized more by qualitative activities and geometry than by intrinsic appearance.The domain models for scene elements are not learned from a corpus ofvideo, but instead, naturally elicited by humans, and represented as probabilistic logic rules within a Markov Logic Network framework. Human elicited models, however, represent object interactions as they occur inthe 3D world rather than describing their appearance projection in some specific 2D image plane. We bridge this gap by recovering qualitative scene geometry to analyze object interactions in the 3D world and thenreasoning about scene geometry, occlusions and common sense domain knowledge using a set of meta-rules. The effectiveness of this approach is demonstrated on a set of videos of public spaces.

Keywords: Scene Understanding, Markov Logic Networks.

1

****

A Data-Driven Approach for Event Prediction

Jenny Yuen and Antonio Torralba
CSAIL MIT
{jenny,torralba }@csail.mit.edu

Abstract. When given a single static picture, humans can not only interpret the instantaneous content captured by the image, but also they are able to infer the chain of dynamic events that are likely to happen inthe near future. Similarly, when a human observes a short video, it is easy to decide if the event taking place in the video is normal or unexpected, even if the video depicts a an unfamiliar place for the viewer. This isin contrast with work in surveillance and outlier event detection, where the models rely on thousands of hours of video recorded at a single place in order to identify what constitutes an unusual event. In this work wepresent a simple method to identify videos with unusual events in a large collection of short video clips. The algorithm is inspired by recent approaches in computer vision that rely on large databases. In this workwe show how, relying on large collections of videos, we can retrieve other

videos similar to the query to build a simple model of the distribution
of expected motions for the query. Consequently, the model can evaluatehow unusual is the video as well as make event predictions. We show how
a very simple retrieval model is able to provide reliable results.
1
****

Activities as Time Series of Human Postures

William Brendel and Sinisa Todorovic
Oregon State University,
Kelley Engineering Cente r, Corvallis , OR 97331, USA
brendelw@onid.orst.edu,sinisa@eecs.oregonstate.edu

Abstract. This paper presents an exemplar-based approach to detecting and localizing human actions, such as running, cycling, and swinging, in realistic videoswith dynamic backgrounds. We show that such activities can be compactly represented as time series of a few snapshots of human-body parts in their most discriminative postures, relative to other activity classes. This enables our approachto efficiently store multiple diverse exemplars per activity class, and quickly retrieve exemplars that best match the query by aligning their short time-series representations. Given a set of example videos of all activity classes, we extractmultiscale regions from all their frames, and then learn a sparse dictionary of most discriminative regions. The Viterbi algorithm is then used to track detections of the learned codewords across frames of each video, resulting in theircompact time-series representations. Dictionary learning is cast within the large-margin framework, wherein we study the effects of /lscript
1and/lscript2regularization on the
sparseness of the resulting dictionaries. Our experiments demonstrate robustness and scalability of our approach on challenging YouTube videos.
1
****

Fast Approximate Nearest Neighbor Methods
for Non-Euclidean Manifolds with Applications
to Human Activity Analysis in Videos

Rizwan Chaudhry1,■and Yuri Ivanov2
1Center for Imaging Science, Johns Hopkins University
3400 N Charles St, Baltimore, MD 21218, USA
rizwanch@cis.jhu.edu
2Mitsubishi Electric Research Laboratories
201 Broadway, Cambridge, MA 02139, USA
yivanov@merl.com

Abstract. Approximate Nearest Neighbor (ANN) methods such as Locality Sensitive Hashing, Semantic Hashing, and Spectral Hashing, provide computationally efficient procedures for finding objects similar to a query object in large datasets. These methods have been successfully applied to search web-scale datasets that can contain millions of images. Unfortunately, the key assumption in these procedures is that objects in the dataset lie in a Euclidean space. This assumption is not always valid and poses a challenge for several computer vision applications where data commonly lies in complex non-Euclidean manifolds. In particular, dynamic data such as human activities are commonly represented as distributions over bags of video words or as dynamical systems. In this paper, we propose two new algorithms that extend Spectral Hashing to non-Euclidean spaces. The first method considers the Riemannian geometry of the manifold and performs Spectral Hashing in the tangent space of the manifold at several points. The second method divides the data into subsets and takes advantage of the kernel trick to perform non-Euclidean Spectral Hashing. For a data set of Nsamples the proposed methods are able to retrieve similar objects in as low as O(K)timecomplexity, where Kis the number of clusters in the data. Since K/lessmuchN,our methods are extremely efficient. We test and evaluate our methods on synthetic data generated from the Unit Hypersphere and the Grassmann manifold. Finally, we show promising results on a human action database.

Keywords: Approximate Nearest Neighbors, Hashing, Non-Euclidean Manifolds, Activity Analysis in Videos.

1

****

The Quadratic-Chi Histogram Distance Family

Ofir Pele and Michael Werman
School of Computer Science
The Hebrew University of Jerusalem
{ofirpele,werman }@cs.huji.ac.il

Abstract. We present a new histogram distance family, the Quadratic-Chi (QC). QC members are Quadratic-Form distances with a cross-bin χ2-like normalization. The cross-bin χ2-like normalization reduces the effect of large bins having undo influence. Normalization was shown to be helpful in many cases, where the χ2histogram distance outperformed the L2norm. However, χ2is sensitive to quantization effects, such as caused by light changes, shape deformations etc. The Quadratic-Form part of QC members takes care of cross-bin relationships (e.g. red and orange), alleviating the quantization problem. We present two new cross-bin histogram distance properties: Similarity-Matrix-Quantization-Invariance andSparseness-Invariance and show that QC distances have these properties. We also show that experimentally they boost performance. QC distances computation time complexity is linear in the number of non-zero entries in the bin-similarity matrix and histograms and it can easily be parallelized. We present results for image retrieval using the Scale Invariant Feature Transform (SIFT) and color image descriptors. In addition, we present results for shape classification using Shape Context (SC) and Inner Distance Shape Context (IDSC). We show that the new QC members outperform state of the art distances for these tasks, while having a short running time. The experimental results show that both the cross-bin property and the normalization are important.

1

****

Membrane Nonrigid Image Registration

Geoffrey Oxholm and Ko Nishino
Department of Computer Science
Drexel University
Philadelphia, PA

Abstract. We introduce a novel nonrigid 2D image registration method that establishes dense and accurate correspondences across images with-out the need of any manual intervention. Our key insight is to model the image as a membrane, i.e., a thin 3D surface, and to constrain its deformation based on its geometric properties. To do so, we derive anovel Bayesian formulation. We impose priors on the moving membrane which act to preserve its shape as it deforms to meet the target. We derive these as curvature weighted first and s econd order derivatives that correspond to the changes in stretching and bending potential energies of themembrane and estimate the registration as the maximum a posteriori. Experimental results on real data demonstrate the effectiveness of our method, in particular, its robustness to local minima and its ability toestablish accurate correspondences ac ross the entire image. The results clearly show that our method overcomes the shortcomings of previous intensity-based and feature-based approaches with conventional uniformsmoothing or diffeomorphic constra ints that suffer from large errors in textureless regions and in areas in-between specified features.

1

****

Affine Puzzle: Realigning Deformed Object Fragments without Correspondences■

Csaba Domokos and Zoltan Kato
Department of Image Processing and Computer Graphics,
University of Szeged
H-6701 Szeged, PO. Box 652., Hungary
Fax: +36 62 546-397
{dcs,kato }@inf.u-szeged.hu

Abstract. This paper is addressing the problem of realigning broken objects without correspondences. We consider linear transformations between the object fragments and present the method through 2D and 3D affine transformations. The basic idea is to construct and solve a polynomial system of equations which provides the unknown parameters of the alignment. We have quantitatively evaluated the proposed algorithm on a large synthetic dataset containing 2D and 3D images. The results show that the method performs well and robust against segmentation errors. We also present experiments on 2D real images as well as on volumetric medical images applied to surgical planning.

1

****

Location Recognition Using Prioritized Feature Matching■

Yunpeng Li, Noah Snavely, and Daniel P. Huttenlocher
Department of Computer Science, Cornell University, Ithaca, NY 14853
{yuli,snavely,dph }@cs.cornell.edu

Abstract. We present a fast, simple location recognition and image localization method that leverages feature correspondence and geometry estimated from large Internet photo collections. Such recovered structure contains a significant amount of useful information about images and image features that is not available when considering images in isolation. For instance, we can predict which views will be the most common, which feature points in a scene are most reliable, and which features in the scene tend to co-occur in the same image. Based on this information, we devise an adaptive, prioritized algorithm for matching a representative set of SIFT features covering a large scene to a query image for efficient localization. Our approach is based on considering features in the scene database, and matching them to query image features, as opposed to more conventional methods that match image features to visual words or database features. We find this approach results in improved performance, due to the richer knowledge of characteristics of the database features comp ared to query image features. We present experiments on two large city-scale photo collections, showing that our algorithm compares favorably to image retrieval-style approaches to location recognition.

Keywords: Location recognition, image registration, image matching, structure from motion.

1

****