1st Annual Catalan Meeting on Computer Vision – (ACMCV)

Abstracts

Centre de Visió per Computador
Bellaterra, Catalonia (Spain)
10 September 2014

ÍNDEX

I. ACMCV 2014 - Abstracts

II. Master Thesis Dissertations - Abstracts

I. ACMCV 2014

Abstracts

Bayesian perspective for the registration of multiple 3D views.

X. Mateo; X. Orriols; X. Binefa.

In CVIU, 2013. From UPF and UAB.

The registration of multiple 3D structures in order to obtain a full-side representation of a scene is a long-time studied subject. Even if the multiple pairwise registrations are almost correct, usually the concatenation of them along a cycle produces a non-satisfactory result at the end of the process due to the accumulation of the small errors. Obviously, the situation can still be worse if, in addition, we have incorrect pairwise correspondences between the views. In this paper, we embed the problem of global multiple views registration into a Bayesian framework, by means of an Expectation–Maximization (EM) algorithm, where pairwise correspondences are treated as missing data and, therefore, inferred through a maximum a posteriori (MAP) process. The presented formulation simultaneously considers uncertainty on pairwise correspondences and noise, allowing a final result which outperforms, in terms of accuracy and robustness, other state-of-the-art algorithms. Experimental results show a reliability analysis of the presented algorithm with respect to the percentage of a priori incorrect correspondences and their consequent effect on the global registration estimation. This analysis compares current state-of-the-art global registration methods with our formulation revealing that the introduction of a Bayesian formulation allows reaching configurations with a lower minimum of the global cost function.

Very Fast Solution to the PnP Problem with Algebraic Outlier Rejection.

L. Ferraz; X. Binefa; F. Moreno-Noguer.

In CVPR, 2014. From UPF/SEPE

We propose a real-time, robust to outliers and accurate solution to the Perspective-n-Point (PnP) problem. The main advantages of our solution are twofold: first, it integrates the outlier rejection within the pose estimation pipeline with a negligible computational overhead; and second, its scalability to arbitrarily large number of correspondences. Given a set of 3D-to-2D matches, we formulate pose estimation problem as a low-rank homogeneous system where the solution lies on its 1D null space. Outlier correspondences are those rows of the linear system which perturb the null space and are progressively detected by projecting them on an iteratively estimated solution of the null space. Since our outlier removal process is based on an algebraic criterion which does not require computing the full-pose and reprojecting back all 3D points on the image plane at each step, we achieve speed gains of more than 100_ compared to RANSAC strategies. An extensive experimental evaluation will show that our solution yields accurate results in situations with up to 50% of outliers, and can process more than 1000 correspondences in less than 5ms.

Leveraging Feature Uncertainty in the PnP Problem.

L. Ferraz; X. Binefa; F. Moreno-Noguer.

In BMVC, 2014. From UPF/SEPE

We propose a real-time and accurate solution to the Perspective-n-Point (PnP) problem – estimating the pose of a calibrated camera from n 3D-to-2D point correspondences– that exploits the fact that in practice the 2D position of not all 2D features is estimated with the same accuracy. Assuming a model of such feature uncertainties is known in advance, we reformulate the PnP problem as a maximum likelihood minimization approximated by an unconstrained Sampson error function, which naturally penalizes the most noisy correspondences. The advantages of this approach are clearly demonstrated in synthetic experiments where feature uncertainties are exactly known.

Pre-estimating the features uncertainties in real experiments is, though, not easy. In this paper we model feature uncertainty as 2D Gaussian distributions representing the sensitivity of the 2D feature detectors to different camera viewpoints. When using these noise models with our PnP formulation we still obtain promising pose estimation results that outperform the most recent approaches.

Compensating inaccurate annotations to train 3D facial landmark localization models.

F. M. Sukno, J. L. Waddington, P. F. Whelan.

In Automatic Face and Gesture Recognition, 2013. From UPF

In this paper we investigate the impact of inconsistency in manual annotations when they are used to train automatic models for 3D facial landmark localization. We start by showing that it is possible to objectively measure the consistency of annotations in a database, provided that it contains replicates (i.e. repeated scans from the same person). Applying such measure to the widely used FRGC database we find that manual annotations currently available are suboptimal and can strongly impair the accuracy of automatic models learnt therefrom. To address this issue, we present a simple algorithm to automatically correct a set of annotations and show that it can help to significantly improve the accuracy of the models in terms of landmark localization errors. This improvement is observed even when errors are measured with respect to the original (not corrected) annotations. However, we also show that if errors are computed against an alternative set of manual annotations with higher consistency, the accuracy of the models constructed using the corrections from the presented algorithm tends to converge to the one achieved by building the models on the alternative, more consistent set.

Stochastic Exploration of Ambiguities for Non-Rigid Shape Recovery.

F. Moreno-Noguer, P. Fua.

In T-PAMI, 2013. From IRI (CSIC-UPC)

Recovering the 3D shape of deformable surfaces from single images is known to be a highly ambiguous problem because many different shapes may have very similar projections. This is commonly addressed by restricting the set of possible shapes to linear combinations of deformation modes and by imposing additional geometric constraints. Unfortunately, because image measurements are noisy, such constraints do not always guarantee that the correct shape will be recovered. To overcome this limitation, we introduce a stochastic sampling approach to efficiently explore the set of solutions of an objective function based on point correspondences. This allows us to propose a small set of ambiguous candidate 3D shapes and then use additional image information to choose the best one.

As a proof of concept, we use either motion or shading cues to this end and show that we can handle a complex objective function without having to solve a difficult nonlinear minimization problem. The advantages of our method are demonstrated on a variety of problems including both real and synthetic data.

A Recurrent Neural Network Approach for 3D Vision-Based Force Estimation.

A. I. Aviles; A. Marban; P. Sobrevilla; J. Fernandez and A. Casals.

In IEEE International Conference on Image Processing Theory, Tools and Applications, 2014.

From UPC

Robotic-assisted minimally invasive surgery has demonstrated its benefits in comparison with traditional procedures. However, one of the major drawbacks of current robotic system approaches is the lack of force feedback. Apart from space restrictions, the main problems of using force sensors are their high cost and the biocompatibility. In this work a proposal based on Vision Based Force Measurement is presented, in which the deformation mapping of the tissue is obtained using the l2-Regularized Optimization class, and the force is estimated via a recurrent neural network that has as inputs the kinematic variables and the deformation mapping. Moreover, the capability of RNN for predicting time series is used in order to deal with tool occlusions. The highlights of this proposal, according to the results, are: knowledge of material properties are not necessary, there is no need of adding extra sensors and a good trade-off between accuracy and efficiency has been achieved.

Fast and Robust ℓ1-averaging-based Pose Estimation for Driving Scenarios.

J. Guerrero; A. D. Sappa; D. Ponsa; A. M. López.

In BMVC, 2013. From CVC, UAB

Robust visual pose estimation is at the core of many computer vision applications, being fundamental for Visual SLAM and Visual Odometry problems. During the last decades, many approaches have been proposed to solve these problems, being RANSAC one of the most accepted and used. However, with the arrival of new challenges, such as large driving scenarios for autonomous vehicles, along with the improvements in the data gathering frameworks, new issues must be considered. One of these issues is the capability of a technique to deal with very large amounts of data while meeting the realtime constraint. With this purpose in mind, we present a novel technique for the problem of robust camera-pose estimation that is more suitable for dealing with large amount of data, which additionally, helps improving the results. The method is based on a combination of a very fast coarse-evaluation function and a robust ℓ1-averaging procedure. Such scheme leads to high-quality results while taking considerably less time than RANSAC.

Experimental results on the challenging KITTI Vision Benchmark Suite are provided, showing the validity of the proposed approach.

A Joint Model for 2D and 3D Pose Estimation from a Single Image.

E. Simo-Serra; A. Quattoni; C. Torras; F. Moreno-Noguer.

In CVPR, 2013. From IRI (CSIC-UPC)

We introduce a novel approach to automatically recover 3D human pose from a single image. Most previous work follows a pipelined approach: initially, a set of 2D features such as edges, joints or silhouettes are detected in the image, and then these observations are used to infer the 3D pose. Solving these two problems separately may lead to erroneous 3D poses when the feature detector has performed poorly. In this paper, we address this issue by jointly solving both the 2D detection and the 3D inference problems. For this purpose, we propose a Bayesian framework that integrates a generative model based on latent variables and discriminative 2D part detectors based on HOGs, and perform inference using evolutionary algorithms. Real experimentation demonstrates competitive results, and the ability of our methodology to provide accurate 2D and 3D pose estimations even when the 2D detectors are inaccurate.

Learning RGB-D Descriptors of Garment Parts for Informed Robot Grasping.

A. Ramisa; G. Alenyà; F. Moreno-Noguer; C. Torras.

In Engineering Applications of Artificial Intelligence, 2014. From IRI (CSIC-UPC)

Robotic handling of textile objects in household environments is an emerging application that has recently received considerable attention thanks to the development of domestic robots. Most current approaches follow a multiple re-grasp strategy for this purpose, in which clothes are sequentially grasped from different points until one of them yields a desired configuration.

In this work we propose a vision-based method, built on the Bag of Visual Words approach that combines appearance and 3D information to detect parts suitable for grasping in clothes, even when they are highly wrinkled.

We also contribute a new, annotated, garment part dataset that can be used for benchmarking classification, part detection, and segmentation algorithms. The dataset is used to evaluate our approach and several state-of-the-art 3D descriptors for the task of garment part detection. Results indicate that appearance is a reliable source of information, but that augmenting it with 3D information can help the method perform better with new clothing items.

Unrolling loopy top-down semantic feedback in convolutional deep networks.

C. Gatta; A. Romero; J. Van de Weijer.

In CVPR Workshops, 2014. From CVC, UAB

In this paper, we propose a novel way to perform topdown semantic feedback in convolutional deep networks for efficient and accurate image parsing. We also show how to add global appearance/semantic features, which have shown to improve image parsing performance in state-of-the-art methods, and was not present in previous convolutional approaches. The proposed method is characterized by an efficient training and a sufficiently fast testing. We use the well known SIFTflow dataset to numerically show the advantages provided by our contributions, and to compare with state-of-the-art image parsing convolutional based approaches.

Stacked Sequential Scale-Space Taylor Context.

C. Gatta; F. Ciompi.

In T-PAMI, 2014. From CVC, UAB

We analyze sequential image labeling methods that sample the posterior label field in order to gather contextual information. We propose an effective method that extracts local Taylor coefficients from the posterior at different scales. Results show that our proposal outperforms state-of-the-art methods on MSRC-21, CAMVID, eTRIMS8 and KAIST2 data sets.

Toward Real-Time Pedestrian Detection Based on a Deformable Template Model.

M. Pedersoli; J. Gonzàlez; X. Hu; X. Roca.

In T-ITS, 2013. From CVC, UAB

Most advanced driving assistance systems already include pedestrian detection systems. Unfortunately, there is still a tradeoff between precision and real time. For a reliable detection, excellent precision-recall such a tradeoff is needed to detect as many pedestrians as possible while, at the same time, avoiding too many false alarms; in addition, a very fast computation is needed for fast reactions to dangerous situations. Recently, novel approaches based on deformable templates have been proposed since these show a reasonable detection performance although they are computationally too expensive for real-time performance. In this paper, we present a system for pedestrian detection based on a hierarchical multiresolution part-based model. The proposed system is able to achieve state-of-the-art detection accuracy due to the local deformations of the parts while exhibiting a speedup of more than one order of magnitude due to a fast coarse-to-fine inference technique. Moreover, our system explicitly infers the level of resolution available so that the detection of small examples is feasible with a very reduced computational cost. We conclude this contribution by presenting how a graphics processing unit-optimized implementation of our proposed system is suitable for real-time pedestrian detection in terms of both accuracy and speed.

Word Spotting and Recognition with Embedded Attributes.

J. Almazán; A. Gordo; A. Fornés; E. Valveny.

In T-PAMI, 2014. From CVC, UAB

This article addresses the problems of word spotting and word recognition on images. In word spotting, the goal is to find all instances of a query word in a dataset of images. In recognition, the goal is to recognize the content of the word image, usually aided by a dictionary or lexicon. We describe an approach in which both word images and text strings are embedded in a common vectorial subspace. This is achieved by a combination of label embedding and attributes learning, and a common subspace regression. In this subspace, images and strings that represent the same word are close together, allowing one to cast recognition and retrieval tasks as a nearest neighbor problem. Contrary to most other existing methods, our representation has a fixed length, is low dimensional, and is very fast to compute and, especially, to compare. We test our approach on four public datasets of both handwritten documents and natural images showing results comparable or better than the state-of-the-art on spotting and recognition tasks.

Coloring Action Recognition in Still Images.

F. S. Khan, R. M. Anwer, J. van de Weijer, A. D. Bagdanov, A. M. López, M. Felsberg.

In IJCV, 2014. From CVC, UAB

In this article we investigate the problem of human action recognition in static images. By action recognition we intend a class of problems which includes both action classification and action detection (i.e. simultaneous localization and classification). Bag-of-words image representations yield promising results for action classification, and deformable part models perform very well object detection. The representations for action recognition typically use only shape cues and ignore color information. Inspired by the recent success of color in image classification and object detection, we investigate the potential of color for action classification and detection in static images. We perform a comprehensive evaluation of color descriptors and fusion approaches for action recognition. Experiments were conducted on the three datasets most used for benchmarking action recognition in still images: Willow, PASCAL VOC 2010 and Stanford-40. Our experiments demonstrate that incorporating color information considerably improves recognition performance, and that a descriptor based on color names outperforms pure color descriptors. Our experiments demonstrate that late fusion of color and shape information outperforms other approaches on action recognition. Finally, we show that the different color–shape fusion approaches result in complementary information and combining them yields state-of-the-art performance for action classification.

RMC-MIL facial behavior categorization.

A. Ruiz, J. Van de Weijer, X. Binefa.

In BMVC, 2014. From UPF

In this work, we address the problem of estimating high-level semantic labels for videos of recorded people by means of analysing their facial expressions. This problem, to which we refer as facial behavior categorization, is a weakly-supervised learning problem where we do not have access to frame-by-frame facial gesture annotations but only weak-labels at the video level are available. Therefore, the goal is to learn a set of discriminative expressions appearing during the training videos and how they determine these labels. Facial behavior categorization can be posed as a Multi-Instance-Learning (MIL) problem and we propose a novel MIL method called Regularized Multi-Concept MIL to solve it. In contrast to previous approaches applied in facial behavior analysis, RMC-MIL follows a Multi-Concept assumption which allows different facial expressions (concepts) to contribute differently to the video-label. Moreover, to handle with the high-dimensional nature of facial-descriptors, RMC-MIL uses a discriminative approach to model the concepts and structured sparsity regularization to discard non-informative features. RMC-MIL is posed as a convex-constrained optimization problem where all the parameters are jointly learned using the Projected-Quasi-Newton method. In our experiments, we use two public data-sets to show the advantages of the Regularized Multi- Concept approach and its improvement compared to existing MIL methods. RMC-MIL outperforms state-of-the-art results in the UNBC data-set for pain detection.

Improving retrieval accuracy of Hierarchical Cellular Trees for generic metric spaces.

C. Ventura; V. Vilaplana; X. Giró-i-Nieto; F. Marqués.

In Multimedia Tools and Applications, 2013. From UPC

Metric Access Methods (MAMs) are indexing techniques which allow working in generic metric spaces. Therefore, MAMs are specially useful for Content-Based Image Retrieval systems based on features which use non L_p norms as similarity measures. MAMs naturally allow the design of image browsers due to their inherent hierarchical structure. The Hierarchical Cellular Tree (HCT), a MAM-based indexing technique, provides the starting point of our work. In this paper, we describe some limitations detected in the original formulation of the HCT and propose some modifications to both the index building and the search algorithm. First, the covering radius, which is defined as the distance from the representative to the furthest element in a node, may not cover all the elements belonging to the node's subtree. Therefore, we propose to redefine the covering radius as the distance from the representative to the furthest element in the node's subtree. This new definition is essential to guarantee a correct construction of the HCT. Second, the proposed Progressive Query retrieval scheme can be redesigned to perform the nearest neighbor operation in a more efficient way. We propose a new retrieval scheme which takes advantage of the benefits of the search algorithm used in the index building. Furthermore, while the evaluation of the HCT in the original work was only subjective, we propose an objective evaluation based on two aspects which are crucial in any approximate search algorithm: the retrieval time and the retrieval accuracy. Finally, we illustrate the usefulness of the proposal by presenting some actual applications.

Multiscale Combinatorial Grouping.

P. Arbelaez, J. Pont-Tuset, J. Barron, F. Marques, J. Malik.

In CVPR, 2014. From UPC

We propose a unified approach for bottom-up hierarchical image segmentation and object candidate generation for recognition, called Multiscale Combinatorial Grouping (MCG). For this purpose, we first develop a fast normalized cuts algorithm. We then propose a high-performance hierarchical segmenter that makes effective use of multiscale information. Finally, we propose a grouping strategy that combines our multiscale regions into highly-accurate object candidates by exploring efficiently their combinatorial space. We conduct extensive experiments on both the BSDS500 and on the PASCAL 2012 segmentation datasets, showing that MCG produces state-of-the-art contours, hierarchical regions and object candidates.

Domain Adaptation of Deformable Part-Based Models.

J. Xu, S. Ramos, D. Vázquez, A. M. López.

In T-PAMI, 2014. From CVC-UAB

The accuracy of object classifiers can significantly drop when the training data (source domain) and the application scenario (target domain) have inherent differences. Therefore, adapting the classifiers to the scenario in which they must operate is of paramount importance. We present novel domain adaptation (DA) methods for object detection. As proof of concept, we focus on adapting the state-of-the-art deformable part-based model (DPM) for pedestrian detection. We introduce an adaptive structural SVM (A-SSVM) that adapts a pre-learned classifier between different domains. By taking into account the inherent structure in feature space (e.g., the parts in a DPM), we propose a structure-aware A-SSVM (SA-SSVM). Neither A-SSVM nor SA-SSVM needs to revisit the source-domain training data to perform the adaptation. Rather, a low number of target-domain training examples (e.g., pedestrians) are used. To address the scenario where there are no target-domain annotated samples, we propose a self-adaptive DPM based on a self-paced learning (SPL) strategy and a Gaussian Process Regression (GPR). Two types of adaptation tasks are assessed: from both synthetic pedestrians and general persons (PASCAL VOC) to pedestrians imaged from an on-board camera. Results show that our proposals avoid accuracy drops as high as 15 points when comparing adapted and non-adapted detectors.

Denoising an Image by Denoising its Components in a Moving Frame.

G. Ghimpeteanu, T. Batard, M. Bertalmío, S. Levine.

In ICISP-Best paper award, 2014. From UPF

In this paper, we provide a new non-local method for image denoising. The key idea we develop is to denoise the components of the image in a well-chosen moving frame instead of the image itself. We prove the relevance of our approach by showing that the PSNR of a grayscale noisy image is lower than the PSNR of its components. Experiments show that applying the Non Local Means algorithms of Buades et al. [5] on the components provides better results than applying it directly on the image.

Diffusion Maps for Multimodal Registration.

G. Piella.

In Sensors, 2014. From UPF

Multimodal image registration is a difficult task, due to the significant intensity variations between the images. A common approach is to use sophisticated similarity measures, such as mutual information, that are robust to those intensity variations. However, these similarity measures are computationally expensive and, moreover, often fail to capture the geometry and the associated dynamics linked with the images. Another approach is the transformation of the images into a common space where modalities can be directly compared. Within this approach, we propose to register multimodal images by using diffusion maps to describe the geometric and spectral properties of the data. Through diffusion maps, the multimodal data is transformed into a new set of canonical coordinates that reflect its geometry uniformly across modalities, so that meaningful correspondences can be established between them. Images in this new representation can then be registered using a simple Euclidean distance as a similarity measure. Registration accuracy was evaluated on both real and simulated brain images with known ground-truth for both rigid and non-rigid registration. Results showed that the proposed approach achieved higher accuracy than the conventional approach using mutual information.

Bayesian View Synthesis and Image-Based Rendering Principles.

S. Pujades, F. Devernay, B. Goldluecke.

In CVPR, 2014. From INRIA Grenoble

In this paper, we address the problem of synthesizing novel views from a set of input images. State of the art methods, such as the Unstructured Lumigraph [4], have been using heuristics to combine information from the original views, often using an explicit or implicit approximation of the scene geometry. While the proposed heuristics have been largely explored and proven to work effectively, a Bayesian formulation was recently introduced [28], formalizing some of the previously proposed heuristics, pointing out which physical phenomena could lie behind each. However, some important heuristics were still not taken into account and lack proper formalization.

We contribute a new physics-based generative model and the corresponding Maximum a Posteriori estimate, providing the desired unification between heuristics- based methods and a Bayesian formulation. The key point is to systematically consider the error induced by the uncertainty in the geometric proxy. We provide an extensive discussion, analyzing how the obtained equations explain the heuristics developed in previous methods. Furthermore, we show that our novel Bayesian model significantly improves the quality of novel views, in particular if the scene geometry estimate is inaccurate.

Color Stabilization Along Time and Across Shots of the Same Scene, for One or Several Cameras of Unknown Specifications.

J. Vazquez-Corral, M. Bertalmío.

In TIP, 2014. From UPF

We propose a method for color stabilization of shots of the same scene, taken under the same illumination, where one image is chosen as reference and one or several other images are modified so that their colors match those of the reference. We make use of two crucial but often overlooked observations: firstly, that the core of the color correction chain in a digital camera is simply a multiplication by a 3x3 matrix; secondly, that to color-match a source image to a reference image we don't need to compute their two color correction matrices, it's enough to compute the operation that transforms one matrix into the other. This operation is a 3x3 matrix as well, which we call H. Once we have H, we just multiply by it each pixel value of the source and obtain an image which matches in color the reference. To compute H we only require a set of pixel correspondences, we don't need any information about the cameras used, neither models nor specifications or parameter values. We propose an implementation of our framework which is very simple and fast, and show how it can be successfully employed in a number of situations, comparing favourably with the state of the art. There is a wide range of applications of our technique, both for amateur and professional photography and video: color matching for multi-camera TV broadcasts, color matching for 3D cinema, color stabilization for amateur video, etc.

Gamut Mapping in Cinematography through Perceptually-based Contrast Modification.

S. W. Zamir, J. Vazquez-Corral, M. Bertalmío.

In IEEE Journal on Selected Topics on Signal Processing, 2014. From UPF

Gamut mapping transforms the colors of an input image to the colors of a target device so as to exploit the full potential of the rendering device in terms of color rendition. In this paper we present spatial gamut mapping algorithms that rely on a perceptually-based variational framework. Our algorithms adapt a well-known image energy functional whose minimization leads to image enhancement and contrast modification. We show how by varying the importance of the contrast term in the image functional we are able to perform gamut reduction and gamut extension. We propose an iterative scheme that allows our algorithms to successfully map the colors from the gamut of the original image to a given destination gamut while keeping the perceived colors close to the original image. Both subjective and objective evaluations validate the promising results achieved via our proposed algorithms.

The Photometry of Intrinsic Images.

M. Serra; O. Penacchio; R. Benavente; M. Vanrell; D. Samaras.

In CVPR, 2014. From CVC-UAB

Intrinsic characterization of scenes is often the best way to overcome the illumination variability artifacts that complicate most computer vision problems, from 3D reconstruction to object or material recognition. This paper examines the deficiency of existing intrinsic image models to accurately account for the effects of illuminant color and sensor characteristics in the estimation of intrinsic images and presents a generic framework which incorporates insights from color constancy research to the intrinsic image decomposition problem. The proposed mathematical formulation includes information about the color of the illuminant and the effects of the camera sensors, both of which modify the observed color of the reflectance of the objects in the scene during the acquisition process. By modeling these effects, we get a ”truly intrinsic” reflectance image, which we call absolute reflectance, which is invariant to changes of illuminant or camera sensors. This model allows us to represent a wide range of intrinsic image decompositions depending on the specific assumptions on the geometric properties of the scene configuration and the spectral properties of the light source and the acquisition system, thus unifying previous models in a single general framework. We demonstrate that even partial information about sensors improves significantly the estimated reflectance images, thus making our method applicable for a wide range of sensors. We validate our general intrinsic image framework experimentally with both synthetic data and natural images.

Low-level Spatio-Chromatic Grouping for Saliency Estimation.

N. Murray; M. Vanrell; X. Otazu; C. A. Párraga.

In T-PAMI, 2013. From CVC, UAB

We propose a saliency model termed SIM (Saliency by Induction Mechanisms) which is based on a low-level spatio-chromatic model that has successfully predicted chromatic induction phenomena. In so doing, we hypothesize that the low-level visual mechanisms that enhance or suppress image detail are also responsible for making some image regions more salient. Moreover, SIM adds geometrical grouplets to enhance complex low-level features such as corners, and suppress relatively simpler features such as edges. Since our model has been fitted on psychophysical chromatic induction data, it is largely non-parametric. SIM outperforms state-of-the-art methods in predicting eye-fixations on two datasets and using two metrics.

Towards Automatic Polyp Detection with a Polyp Appearance Model.

J. Bernal, F. J. Sánchez, F. Vilariño.

In PR, 2012. From CVC, UAB

This work aims at automatic polyp detection by using a model of polyp appearance in the context of the analysis of colonoscopy videos. Our method consists of three stages: region segmentation, region description and region classification. The performance of our region segmentation method guarantees that if a polyp is present in the image, it will be exclusively and totally contained in a single region. The output of the algorithm also defines which regions can be considered as non-informative. We define as our region descriptor the novel Sector Accumulation-Depth of Valleys Accumulation (SA-DOVA), which provides a necessary but not sufficient condition for the polyp presence. Finally, we classify our segmented regions according to the maximal values of the SA-DOVA descriptor. Our preliminary classification results are promising, especially when classifying those parts of the image that do not contain a polyp inside.

Detection of wrinkle frames in endoluminal videos using betweenness centrality measures for images.

S. Segui; M. Drozdzal; E. Zaytseva; C. Malagelada; F. Azpiroz; P. Radeva; J. Vitria.

In IEEE Journal of Biomedical and Health Informatics, 2014. From CVC, UAB

Intestinal contractions are one of the most important events to diagnose motility pathologies of the small intestine. When visualized by wireless capsule endoscopy (WCE), the sequence of frames that represents a contraction is characterized by a clear wrinkle structure in the central frames that corresponds to the folding of the intestinal wall. In this paper we present a new method to robustly detect wrinkle frames in full WCE videos by using a new mid-level image descriptor that is based on a centrality measure proposed for graphs. We present an extended validation, carried out in a very large database, that shows that the proposed method achieves state of the art performance for this task.

A framework for optimal kernel-based manifold embeddingof medical image data.

V. A. Zimmer; K. Lekadir; C. Hoogendoorn; A. F. Frangi; G. Piella.

In Computerized Medical Imaging and Graphics, 2014. From UPF

Kernel-based dimensionality reduction is a widely used technique in medical image analysis. To fully unravel the underlying nonlinear manifold the selection of an adequate kernel function and of its free parameters is critical. In practice, however, the kernel function is generally chosen as Gaussian or polynomial and such standard kernels might not always be optimal for a given image dataset or application. In this paper, we present a study on the effect of the kernel functions in nonlinear manifold embedding of medical image data. To this end, we first carry out a literature review on existing advanced kernels developed in the statistics, machine learning, and signal processing communities. In addition, we implement kernel-based formulations of well-known nonlinear dimensional reduction techniques such as Isomap and Locally Linear Embedding, thus obtaining a unified framework for manifold embedding using kernels. Subsequently, we present a method to automatically choose a kernel function and its associated parameters from a pool of kernel candidates, with the aim to generate the most optimal manifold embeddings. Furthermore, we show how the calculated selection measures can be extended to take into account the spatial relationships in images, or used to combine several kernels to further improve the embedding results. Experiments are then carried out on various synthetic and phantom datasets for numerical assessment of the methods. Furthermore, the workflow is applied to real data that include brain manifolds and multispectral images to demonstrate the importance of the kernel selection in the analysis of high-dimensional medical images.

Improved myocardial motion estimation combining tissue Doppler and B-mode echocardiographic images.

A. R. Porras; M. Alessandrini; M. De Craene; N. Duchateau; M. Sitges; B. H. Bijnens; H. Delingette; M. Sermesant; J. D'hooge; A. F. Frangi; G. Piella.

In IEEE Transactions on Medical Imaging, 201. From UPF

We propose a technique for myocardial motion estimation based on image registration using both B-mode echocardiographic images and tissue Doppler sequences acquired interleaved. The velocity field is modeled continuously using B-splines and the spatiotemporal transform is constrained to be diffeomorphic. Images before scan conversion are used to improve the accuracy of the estimation. The similarity measure includes a model of the speckle pattern distribution of B-mode images. It also penalizes the disagreement between tissue Doppler velocities and the estimated velocity field. Registration accuracy is evaluated and compared to other alternatives using a realistic synthetic dataset, obtaining mean displacement errors of about 1 mm. Finally, the method is demonstrated on data acquired from 6 volunteers, both at rest and during exercise. Robustness is tested against low image quality and fast heart rates during exercise. Results show that our method provides a robust motion estimate in these situations.

Evaluation of the Capabilities of Confidence Measures for Assessing Optical Flow Quality.

P. Márquez-Valle, D. Gil, A. Hernàndez-Sabaté.

In ICCV Workshops, 2013. From CVC, UAB

Assessing Optical Flow (OF) quality is essential for its further use in reliable decision support systems. The absence of ground truth in such situations leads to the computation of OF Confidence Measures (CM) obtained from either input or output data. A fair comparison across the capabilities of the different CM for bounding OF error is required in order to choose the best OF-CM pair for discarding points where OF computation is not reliable. This paper presents a statistical probabilistic framework for assessing the quality of a given CM. Our quality measure is given in terms of the percentage of pixels whose OF error bound cannot be determined by CM values. We also provide statistical tools for the computation of CM values that ensures a given accuracy of the flow field.

Region-based particle filter for video object segmentation

D. Varas, F. Marques

In CVPR, 2014. From UPC, Tech

We present a video object segmentation approach that extends the particle filter to a region-based image representation. Image partition is considered part of the particle filter measurement, which enriches the available information and leads to a re-formulation of the particle filter. The prediction step uses a co-clustering between the previous image object partition and a partition of the current one, which allows us to tackle the evolution of non-rigid structures. Particles are defined as unions of regions in the current image partition and their propagation is computed through a single co-clustering. The proposed technique is assessed on the SegTrack dataset, leading to satisfactory perceptual results and obtaining very competitive pixel error rates compared with the state-of-the-art methods

A variational model for gradient-based video editing.

S. Rida, F. Gabriele, A. Pablo, C. Vicent.

In IJCV, 2013. From UPF

In this work we present a gradient-based variational model for video editing, addressing the problem of propagating gradient-domain information along the optical flow of the video. The resulting propagation is temporally consistent and blends seamlessly with its spatial surroundings. In addition, the presented model is able to cope with additive illumination changes and handles occlusions/dis-occlusions. The problem of propagation along the optical flow arises in different video editing applications. In this work we consider the application where a user edits a frame by modifying the texture of an object's surface and wishes to propagate this editing throughout the video.

Constrained optical flow estimation as a matching problem.

M. Mozerov.

In TIP, 2013. From CVC, UAB

In general, discretization in the motion vector domain yields an intractable number of labels. In this paper, we propose an approach that can reduce general optical flow to the constrained matching problem by pre-estimating a 2-D disparity labeling map of the desired discrete motion vector function. One of the goals of the proposed paper is estimating coarse distribution of motion vectors and then utilizing this distribution as global constraints for discrete optical flow estimation. This pre-estimation is done with a simple frame-to-frame correlation technique also known as the digital symmetric-phase-only-filter (SPOF). We discover a strong correlation between the output of the SPOF and the motion vector distribution of the related optical flow. A two step matching paradigm for optical flow estimation is applied: pixel accuracy (integer flow) and subpixel accuracy estimation. The matching problem is solved by global optimization. Experiments on the Middlebury optical flow datasets confirm our intuitive assumptions about strong correlation between motion vector distribution of optical flow and maximal peaks of SPOF outputs. The overall performance of the proposed method is promising and achieves state-of-the-art results on the Middlebury benchmark.

Object segmentation in images using EEG signals.

E. Mohedano; G. Healy; K. McGuinness; X. Giro-i-Nieto; N. O'Connor, A. Smeaton.

In ACM Multimedia, 2014. From UPC

This paper explores the potential of brain-computer interfaces in segmenting objects from images. Our approach is centered around designing an effective method for displaying the image parts to the users such that they generate measurable brain reactions. When an image region, specifically a block of pixels, is displayed we estimate the probability of the block containing the object of interest using a score based on EEG activity. After several such blocks are displayed, the resulting probability map is binarized and combined with the GrabCut algorithm to segment the image into object and background regions. This study shows that BCI and simple EEG analysis are useful in locating object boundaries in images.

A Neurodynamical Model of Brightness Induction in V1.

O. Penacchio, X. Otazu, L. Dempere-Marco.

In PLoS ONE, 2013. From CVC, UAB

Brightness induction is the modulation of the perceived intensity of an area by the luminance of surrounding areas. Recent neurophysiological evidence suggests that brightness information might be explicitly represented in V1, in contrast to the more common assumption that the striate cortex is an area mostly responsive to sensory information. Here we investigate possible neural mechanisms that offer a plausible explanation for such phenomenon. To this end, a neurodynamical model which is based on neurophysiological evidence and focuses on the part of V1 responsible for contextual influences is presented. The proposed computational model successfully accounts for well known psychophysical effects for static contexts and also for brightness induction in dynamic contexts defined by modulating the luminance of surrounding areas. This work suggests that intra-cortical interactions in V1 could, at least partially, explain brightness induction effects and reveals how a common general architecture may account for several different fundamental processes, such as visual saliency and brightness induction, which emerge early in the visual processing pathway.

The Perceived Quality of Undistorted Natural Images.

D. Kane, M. Bertalmío.

In VSS, 2014. From UPF

We investigate the perceived quality of natural images. To do so, we linearly scale the luminance range of high dynamic range images to generate a set of tone-mapped images that cover the full range of mean-luminance and contrast values that a CRT monitor can display. Image patches are displayed on a uniform black, grey or white background and subjects are asked to evaluate the quality of each image on a 0-9 scale. We find that image quality scores can be predicted using a three-stage model: First, luminance is converted to lightness using an expansive power-law that varies with the background luminance (γblack=0.3, γgray=0.35, γwhite=0.45). Second, the standard deviation of the gamma-adjusted, lightness image is computed. Third, the standard deviation is passed through an expansive power-law (γ=0.3) to estimate the perceived contrast of an image. This metric can accurately predict the average image quality scores over the full range of onscreen luminance and contrast values investigated (r=0.94, p<0.0001). A second investigation reveals that the proposed contrast metric is linearly related to image quality scores for all test images with a mean Pearson's correlation of 0.87 (N=128), however the slope of the function varies substantially between images and we are unable to model this effect. The proposed contrast metric is able to predict the perceived quality of tone-mapped images in the database of Cadik et al. (2008) despite the existence of a wide variety of image artefacts (ringing, colour distortions, ect) in the image set (r=0.85, p<0.0001). Finally, we note that the proposed super-threshold contrast metric performs histogram equalisation on the luminance distribution and removes skew from the contrast distribution of natural scenes, suggesting an optimal coding strategy.

Which tone-mapping is the best? A comparative study of tone-mapping perceived quality.

X. Cerdá; C. A. Parraga; X. Otazu.

In Perception, 2014. From CVC, UAB

High-dynamic-range (HDR) imaging refers to the methods designed to increase the brightness dynamic range present in standard digital imaging techniques. This increase is achieved by taking the same picture under different exposure values and mapping the intensity levels into a single image by way of a tone-mapping operator (TMO). Currently, there is no agreement on how to evaluate the quality of different TMOs. In this work we psychophysically evaluate 15 different TMOs obtaining rankings based on the perceived properties of the resulting tone-mapped images. We performed two different experiments on a CRT calibrated display using 10 subjects: (1) a study of the internal relationships between grey-levels and (2) a pairwise comparison of the resulting 15 tone-mapped images. In (1) observers internally matched the grey-levels to a reference inside the tone-mapped images and in the real scene. In (2) observers performed a pairwise comparison of the tone-mapped images alongside the real scene. We obtained two rankings of the TMOs according their performance. In (1) the best algorithm was ICAM by J.Kuang et al (2007) and in (2) the best algorithm was a TMO by Krawczyk et al (2005). Our results also show no correlation between these two rankings. [CAP and XO are supported by grant TIN2010-21771-C02-01 from the Spanish Ministry of Science.]

6 Seconds of Sound and Vision: Creativity in Micro-Videos.

M. Redi, N. O'Hare, R. Schifanella, M. Trevisiol, A. Jaimes.

In CVPR, 2014. From Yahoo Labs

The notion of creativity, as opposed to related concepts such as beauty or interestingness, has not been studied from the perspective of automatic analysis of multimedia content. Meanwhile, short online videos shared on social media platforms, or micro-videos, have arisen as a new medium for creative expression. In this paper we study creative microvideos in an effort to understand the features that make a video creative, and to address the problem of automatic detection of creative content. Defining creative videos as those that are novel and have aesthetic value, we conduct a crowdsourcing experiment to create a dataset of over 3,800 microvideos labelled as creative and non-creative. We propose a set of computational features that we map to the components of our definition of creativity, and conduct an analysis to determine which of these features correlate most with creative video. Finally, we evaluate a supervised approach to automatically detect creative video, with promising results, showing that it is necessary to model both aesthetic value and novelty to achieve optimal classification accuracy.

II. Master Thesis Dissertations

Abstracts

2013-2014

Feature selection using synthetic view matching for large-scale image search

Miquel Ferrarons Betrian

Supervisors: Tomasz Adamek (CTO de Catchoom) i Xavier Giró (UPC).

This thesis presents a method to select the most important visual features from images to be indexed for a fast large-scale visual search scenario. The method is based on matching synthetic views of the reference images, and it has been successfully applied to two different scenarios: one for reducing the amount of information stored for the reference images, and another for improving the robustness against viewpoint changes between the query and the reference images.

Performance in trampoline

Carlos Puig Toledo

Supervisors: Jordi Gonzalez (UAB), Sergio Escalera (CVC)

In this master thesis, it is proposed to capture multi-modal RGB-Depth data obtained by a Kinect device, synchronize and align the captured modalities with a frame rate near 30FPS, and use Computer Vision techniques and methods in order to extract a relevant indicator as is the landing point concerning a jump.

Design and Implementation of a Semi-Automatic Bodymarks Detection System

Eduard Ramon Maldonado

Supervisor: Ernest Valveny (UAB)

This master thesis presents a semi-automatic system used to detect concrete points of the body in frontal and lateral views. Afterwards these points are used to obtain 3D information of the body. In order to build the system, all images from database has been normalized in scale and rotation using two points, which are the only input that the system requires. Then, expectation maximization algorithm has been applied to model the location of the rest of the points with respect a reference. Finally, a random forest classifier has been trained to refine the position of each point within its probable region using ORB descriptors.

3D reconstruction and recognition using structured light.

Antonio Esteban Lansaque

Supervisor: Javier Ruiz (UPC)

This work covers the problem of 3D reconstruction, recognition and 6DOF pose estimation using Matlab and Point Cloud Library (PCL). The goal of this project is to reconstruct a 3D scene and to align an object model of the industrial pieces onto the reconstructed scene. Whereas reconstruction algorithm is based on stereo techniques, the recognition algorithm is based on SHOT descriptors computed on a set of uniform keypoints. Correspondences are used to estimate a first 6DOF transformation that maps the model onto the scene and then ICP algorithm is used to refine the transformation.

Experiments, which were used to check the effectiveness of the proposed algorithm, were carried out in a lab environment. Although results are not real time results, the algorithm ends up with high rates of recognition.

Exploiting User Interaction and Object Candidates for Instance Retrieval and Object Segmentation

Amaia Salvador Aguilera

Supervisor: Xavier Giró (UPC)

This thesis addresses two of the main challenges nowadays for computer vision: object segmentation and visual instance retrieval. The methodologies proposed to solve both problems are based on the use of object candidates and human computation in the computer vision loop. In the object segmentation side, this work explores how human computation can be useful to achieve better segmentation results, by combining users' traces with a segmentation algorithm based on object candidates. On the other hand, the instance retrieval problem is also addressed using object candidates to compute local features, and involving the user in the retrieval loop by applying relevance feedback strategies.

Q-Learning in an Open-Space Combat Scenario for Real-Time Strategy Games

Ferran Mestres

Supervisors: Jesús Cerquides and Josep Lluís Arcos

This thesis aims to explore the design of an Artificial Intelligence bot for Real-Time Strategy games. Specifically, the thesis focus on the use of the Q-learning algorithm to train a bot that controls a group of units in the commercial game Starcraft.

The bot learns a set of winning strategies to be applied both at squad and at unit level, understanding a squad as a small group of units.

Experiments are conducted on open-space combats, scenarios that abstract the idea of a combat to death. Moreover, it explores the possibility of an agent with a knowledge base formed after training over different scenarios, and how it will perform in a combat on a more realistic in-game situation.