FINAL YEAR PROJECTS

IMAGE PROCESSING 2016-2017

TITLES WITH ABSTRACTS

CALL US @ 994-479-3398, (0452) 4390702

17 Years of Experience • Automated Services • 24/7 Help Desk Support
Advanced Technologies And Tools • Legitimate Members Of all Journals

f/elysiumtechnologiespvtltd  /elysiumind  in/elysium-technologies-pvt
ELYSIUM TECHNOLOGIES
PRIVATE LIMITED

SINGAPORE | CHENNAI | MADURAI | TRICHY | DINDIGUL | SIVAKASI

CORPORATE OFFICE: MADURAI

#227, Church Road, Annanagar, Madurai - 625020.
(+91)452-4390702, 4390702, 4392702, 4394702.
info@elysiumtechnologies.com, elysiumtechnologies@gmail.com

BRANCH OFFICE: CHENNAI

S.P.Towers, #81 Valluvar Kottam High Road, Nungambakkam, Chennai - 600034. 004-42072702, (+91) 9600354638,
chennai@elysiumtechnologies.com

BRANCH OFFICE: TRICHY

#15, 3rd Floor, SI Towers, Melapudur Main Road, Trichy - 625001.
(+91) 431-4002234, 9790464324
trichy@elysiumtechnologies.com

BRANCH OFFICE: SIVAKASI

#661 A, Yanagiri Road, Cornation Bus Stop, Sivakasi - 626123, (+91)9677789339.
sivakasi@elysiumtechnologies.com

BRANCH OFFICE: DINDIGUL

#23, 100 ft road, opp to KVB Bank (RM colony branch),
West Govindhapuram, RM colony, Dindigul -62400
dindigul@elysiumtechnologies.com
**ETPL DIP - 001**  
Underwater Depth Estimation and Image Restoration Based on Single Images.

In underwater environments, the scattering and absorption phenomena affect the propagation of light, degrading the quality of captured images. In this work, the authors present a method based on a physical model of light propagation that takes into account the most significant effects to image degradation: absorption, scattering, and backscattering. The proposed method uses statistical priors to restore the visual quality of the images acquired in typical underwater scenarios.

**ETPL DIP - 002**  

Clustering for hyperspectral images (HSIs) is a very challenging task due to its inherent complexity. In this paper, we propose a novel spectral-spatial sparse subspace clustering $S^4C$ algorithm for hyperspectral remote sensing images. First, by treating each kind of land-cover class as a subspace, we introduce the sparse subspace clustering (SSC) algorithm to HSIs. Then, considering the spectral and spatial properties of HSIs, the high spectral correlation and rich spatial information of the HSIs are taken into consideration in the SSC model to obtain a more accurate coefficient matrix, which is used to build the adjacent matrix. Finally, spectral clustering is applied to the adjacent matrix to obtain the final clustering result. Several experiments were conducted to illustrate the performance of the proposed $S^4C$ algorithm.

**ETPL DIP - 003**  
Beyond colour Difference: Residual Interpolation for colour Image Demosaicking.

In this paper, we propose residual interpolation (RI) as an alternative to color difference interpolation, which is a widely accepted technique for color image demosaicking. Our proposed RI performs the interpolation in a residual domain, where the residuals are differences between observed and tentatively estimated pixel values. Our hypothesis for the RI is that if image interpolation is performed in a domain with a smaller Laplacian energy, its accuracy is improved. Based on the hypothesis, we estimate the tentative pixel values to minimize the Laplacian energy of the residuals. We incorporate the RI into the gradient-based threshold free algorithm, which is one of the state-of-the-art Bayer demosaicking algorithms. Experimental results demonstrate that our proposed demosaicking algorithm using the RI surpasses the state-of-the-art algorithms for the Kodak, the IMAX, and the beyond Kodak data sets.
Conventional line detection methods are mainly based on the binary edge map. This letter proposes a new line detection method that directly extracts line features from the image edge fields of the synthetic aperture radar (SAR) images. In the proposed method, the strength and direction of each field point are first obtained using a ratio-based edge filter. Then, the accumulation weight of the field point is jointly computed using its strength and direction. The direction of a field point on the line is essentially the orientation of the line. Furthermore, a field point on a strong line should be distinguished from a field point on a weak line. Thus, the accumulation weights of different field points are not equal. By summing up the accumulation weights, the straight lines in the SAR image space are directly converted into several local peaks in the parameter space. A sort-window peak detection method is proposed to suppress the spurious secondary peaks in the parameter space. The experimental results show that the proposed line detection method is robust to noise and has a good antiocclusion ability. The proposed method performs well in terms of true positive detection rate and detection accuracy for both synthetic and real-world images.

---

Image denoising has been a well-studied problem for imaging systems, especially imaging sensors. Despite remarkable progress in the quality of denoising algorithms, persistent challenges remain for a wide class of general images. In this paper, we present a new concept of sequence-to-sequence similarity (SSS). This similarity measure is an efficient method to evaluate the content similarity for images, especially for edge information. The approach differs from the traditional image processing techniques, which rely on pixel and block similarity. Based on this new concept, we introduce a new SSS-based filter for image denoising. The new SSS-based filter utilizes the edge information in the corrupted image to address image denoising problems. We demonstrate the filter by incorporating it into a new SSS-based image denoising algorithm to remove Gaussian noise. Experiments are performed over synthetic and experimental data. The performance of our methodology is experimentally verified on a variety of images and Gaussian noise levels. The results demonstrate that the proposed method's performance exceeds several current state-of-the-art works, which are evaluated both visually and quantitatively. The presented framework opens up new perspectives in the use of SSS methodologies for image processing applications to replace the traditional pixel-to-pixel similarity or block-to-block similarity.
Visual computing technologies have traditionally been developed for conventional setups where air is the surrounding medium for the user, the display, and/or the camera. However, given mankind's increasingly need to rely on the oceans to solve the problems of future generations (such as offshore oil and gas, renewable energies, and marine mineral resources), there is a growing need for mixed-reality applications for use in water. This article highlights the various research challenges when changing the medium from air to water, introduces the concept of underwater mixed environments, and presents recent developments in underwater visual computing applications.

Semi-supervised image classification aims to classify a large quantity of unlabeled images by typically harnessing scarce labeled images. Existing semi-supervised methods often suffer from inadequate classification accuracy when encountering difficult yet critical images, such as outliers, because they treat all unlabeled images equally and conduct classifications in an imperfectly ordered sequence. In this paper, we employ the curriculum learning methodology by investigating the difficulty of classifying every unlabeled image. The reliability and the discriminability of these unlabeled images are particularly investigated for evaluating their difficulty. As a result, an optimized image sequence is generated during the iterative propagations, and the unlabeled images are logically classified from simple to difficult. Furthermore, since images are usually characterized by multiple visual feature descriptors, we associate each kind of features with a teacher, and design a multi-modal curriculum learning (MMCL) strategy to integrate the information from different feature modalities. In each propagation, each teacher analyzes the difficulties of the currently unlabeled images from its own modality viewpoint. A consensus is subsequently reached among all the teachers, determining the currently simplest images (i.e., a curriculum), which are to be reliably classified by the multi-modal learner. This well-organized propagation process leveraging multiple teachers and one learner enables our MMCL to outperform five state-of-the-art methods on eight popular image data sets.
As the unique identification of a vehicle, license plate is a key clue to uncover over-speed vehicles or the ones involved in hit-and-run accidents. However, the snapshot of over-speed vehicle captured by surveillance camera is frequently blurred due to fast motion, which is even unrecognizable by human. Those observed plate images are usually in low resolution and suffer severe loss of edge information, which cast great challenge to existing blind deblurring methods. For license plate image blurring caused by fast motion, the blur kernel can be viewed as linear uniform convolution and parametrically modeled with angle and length. In this paper, we propose a novel scheme based on sparse representation to identify the blur kernel. By analyzing the sparse representation coefficients of the recovered image, we determine the angle of the kernel based on the observation that the recovered image has the most sparse representation when the kernel angle corresponds to the genuine motion angle. Then, we estimate the length of the motion kernel with Radon transform in Fourier domain. Our scheme can well handle large motion blur even when the license plate is unrecognizable by human. We evaluate our approach on real-world images and compare with several popular state-of-the-art blind image deblurring algorithms. Experimental results demonstrate the superiority of our proposed approach in terms of effectiveness and robustness.

This paper presents a technique for transmitting information efficiently and securely, hiding confidential messages on seemingly innocent messages using steganography. The insertion technique in the least significant bit is used to insert images into digital pictures or other secret watermark. Artificial Neural Networks are used in the process of withdrawal of encrypted information acting as keys that determine the existence of hidden information.
Recently, various patch-based approaches have emerged for high and very high resolution multispectral image classification and indexing. This comes as a consequence of the most important particularity of multispectral data: objects are represented using several spectral bands that equally influence the classification process. In this letter, by using a patch-based approach, we are aiming at extracting descriptors that capture both spectral information and structural information. Using both the raw texture data and the high spectral resolution provided by the latest sensors, we propose enhanced image descriptors based on Gabor, spectral histograms, spectral indices, and bag-of-words framework. This approach leads to a scene classification that outperforms the results obtained when employing the initial image features. Experimental results on a WorldView-2 scene and also on a test collection of tiles created using Sentinel 2 data are presented. A detailed assessment of speed and precision was provided in comparison with state-of-the-art techniques. The broad applicability is guaranteed as the performances obtained for the two selected data sets are comparable, facilitating the exploration of previous and newly lunched satellite missions.

This paper addresses the problem of detecting coherent motions in crowd scenes and presents its two applications in crowd scene understanding: semantic region detection and recurrent activity mining. It processes input motion fields (e.g., optical flow fields) and produces a coherent motion field named thermal energy field. The thermal energy field is able to capture both motion correlation among particles and the motion trends of individual particles, which are helpful to discover coherency among them. We further introduce a two-step clustering process to construct stable semantic regions from the extracted time-varying coherent motions. These semantic regions can be used to recognize pre-defined activities in crowd scenes. Finally, we introduce a cluster-and-merge process, which automatically discovers recurrent activities in crowd scenes by clustering and merging the extracted coherent motions. Experiments on various videos demonstrate the effectiveness of our approach.
**ETPL DIP - 012**  
Unsupervised Co-Segmentation for Indefinite Number of Common Foreground Objects.

Co-segmentation addresses the problem of simultaneously extracting the common targets appeared in multiple images. Multiple common targets involved object co-segmentation problem, which is very common in reality, has been a new research hotspot recently. In this paper, an unsupervised object co-segmentation method for indefinite number of common targets is proposed. This method overcomes the inherent limitation of traditional proposal selection-based methods for multiple common targets involved images while retaining their original advantages for objects extracting. For each image, the proposed multi-search strategy extracts each target individually and an adaptive decision criterion is raised to give each candidate a reliable judgment automatically, i.e., target or non-target. The comparison experiments conducted on public data sets iCoseg, MSRC, and a more challenging data set Coseg-INCT demonstrate the superior performance of the proposed method.

**ETPL DIP - 013**  
Hierarchical Discriminative Feature Learning for Hyper spectral Image Classification.

Building effective image representations from hyperspectral data helps to improve the performance for classification. In this letter, we develop a hierarchical discriminative feature learning algorithm for hyperspectral image classification, which is a deformation of the spatial-pyramid-matching model based on the sparse codes learned from the discriminative dictionary in each layer of a two-layer hierarchical scheme. The pooling features achieved by the proposed method are more robust and discriminative for the classification. We evaluate the proposed method on two hyperspectral data sets: Indiana Pines and Salinas scene. The results show our method possessing state-of-the-art classification accuracy.
Extraction of residential areas plays an important role in remote sensing image processing. Extracted results can be applied to various scenarios, including disaster assessment, urban expansion, and environmental change research. Quality residential areas extracted from a remote sensing image must meet three requirements: well-defined boundaries, uniformly highlighted residential area, and no background redundancy in residential areas. Driven by these requirements, this study proposes a global and local saliency analysis model (GLSA) for the extraction of residential areas in high-spatial-resolution remote sensing images. In the proposed model, a global saliency map based on quaternion Fourier transform (QFT) and a global saliency map based on adaptive directional enhancement lifting wavelet transform (ADE-LWT) are generated along with a local saliency map, all of which are fused into a main saliency map based on complementarities. In order to analyze the correlation among spectrums in the remote sensing image, the phase spectrum information of QFT is used on the multispectral images for producing a global saliency map. To acquire the texture and edge features of different scales and orientations, the coefficients acquired by ADE-LWT are used to construct another global saliency map. To discard redundant backgrounds, the amplitude spectrum of the Fourier transform and the spatial relations among patches are introduced into the panchromatic image to generate the local saliency map. Experimental results indicate that the GLSA model can better define the boundaries of residential areas and achieve complete residential areas than current methods. Furthermore, the GLSA model can prevent redundant backgrounds in residential areas and thus acquire more accurate residential areas.

Live fish recognition is one of the most crucial elements of fisheries survey applications where the vast amount of data is rapidly acquired. Different from general scenarios, challenges to underwater image recognition are posted by poor image quality, uncontrolled objects and environment, and difficulty in acquiring representative samples. In addition, most existing feature extraction techniques are hindered from automation due to involving human supervision. Toward this end, we propose an underwater fish recognition framework that consists of a fully unsupervised feature learning technique and an error-resilient classifier. Object parts are initialized based on saliency and relaxation labeling to match object parts correctly. A non-rigid part model is then learned based on fitness, separation, and discrimination criteria. For the classifier, an unsupervised clustering approach generates a binary class hierarchy, where each node is a classifier. To exploit information from ambiguous images, the notion of partial classification is introduced to assign coarse labels by optimizing the benefit of indecision made by the classifier. Experiments show that the proposed framework achieves high accuracy on both public and self-collected underwater fish images with high uncertainty and class imbalance.
**Statistics of Natural Stochastic Textures and Their Application in Image De-noising.**

Natural stochastic textures (NSTs), characterized by their fine details, are prone to corruption by artifacts, introduced during the image acquisition process by the combined effect of blur and noise. While many successful algorithms exist for image restoration and enhancement, the restoration of natural textures and textured images based on suitable statistical models has yet to be further improved. We examine the statistical properties of NST using three image databases. We show that the Gaussian distribution is suitable for many NST, while other natural textures can be properly represented by a model that separates the image into two layers; one of these layers contains the structural elements of smooth areas and edges, while the other contains the statistically Gaussian textural details. Based on these statistical properties, an algorithm for the denoising of natural images containing NST is proposed, using patch-based fractional Brownian motion model and regularization by means of anisotropic diffusion. It is illustrated that this algorithm successfully recovers both missing textural details and structural attributes that characterize natural images. The algorithm is compared with classical as well as the state-of-the-art denoising algorithms.

**Large Polari metric SAR Data Semi-Supervised Classification with Spatial-Anchor Graph.**

Recently, graph-based semi-supervised classification (SSC) has attracted considerable attentions as it could enhance classification accuracy by utilizing only a few labeled samples and large numbers of unlabeled samples via graphs. However, the construction of graphs is time consuming especially for large-scale polarimetric synthetic aperture radar (PolSAR) data. Moreover, speckle noise in images remarkably degrades the accuracy of the constructed graph. To address these two issues, this paper proposes a novel spatial-anchor graph for large-scale PolSAR terrain classification. First, the PolSAR image is segmented to obtain homogeneous regions. The features of each pixel are weighted by that of the surrounding pixels from the homogeneous regions to reduce the influence of the speckle noise. Second, Wishart distance-based clustering is performed on the weighted features, and the cluster centers are computed and serve as initial anchors. Then, the label of each pixel is predicted by the label of its nearest anchors on the spatial-anchor graph which is constructed through solving an optimization problem. Experimental results on synthesized PolSAR data and real ones from different approaches show that the proposed method reduces the computational complexity to a linear time, and the graph combined with the spatial information suppresses the speckle noise and enhances the classification accuracy in comparison with state-of-the-art graph-based SSCs when only a small number of labeled samples are available.
We use methods from Riemann geometry to investigate transformations between the color spaces of color-normal and color-weak observers. The two main applications are the simulation of the perception of a color weak observer for a color-normal observer, and the compensation of color images in a way that a color-weak observer has approximately the same perception as a color-normal observer. The metrics in the color spaces of interest are characterized with the help of ellipsoids defined by the just-noticeable-differences between the colors which are measured with the help of color-matching experiments. The constructed mappings are the isometries of Riemann spaces that preserve the perceived color differences for both observers. Among the two approaches to build such an isometry, we introduce normal coordinates in Riemann spaces as a tool to construct a global color-weak compensation map. Compared with the previously used methods, this method is free from approximation errors due to local linearizations, and it avoids the problem of shifting locations of the origin of the local coordinate system. We analyze the variations of the Riemann metrics for different observers obtained from new color-matching experiments and describe three variations of the basic method. The performance of the methods is evaluated with the help of semantic differential tests.

In this letter, a novel synthetic aperture radar (SAR) image registration method, including two operators for feature detection and arborescence network matching (ANM) for feature matching, is proposed. The two operators, namely, SAR scale-invariant feature transform (SIFT) and R-SIFT, can detect corner points and texture points in SAR images, respectively. This process has an advantage of preserving two types of feature information in SAR images simultaneously. The ANM algorithm has a two-stage process for finding matching pairs. The backbone network and the branch network are successively built. This ANM algorithm combines feature constraints with spatial relations among feature points and possesses a larger number of matching pairs and higher subpixel matching precision than the original version. Experimental results on various SAR images show that the proposed method provides superior performance than other approaches investigated.
In this paper, we explore the potentialities of using wavelet-based multivariate models for the classification of very high resolution optical images. A strategy is proposed to apply these models in a supervised classification framework. This strategy includes a content-based image retrieval analysis applied on a texture database prior to the classification in order to identify which multivariate model performs the best in the context of application. Once identified, the best models are further applied in a supervised classification procedure by extracting texture features from a learning database and from regions obtained by a presegmentation of the image to classify. The classification is then operated according to the decision rules of the chosen classifier. The use of the proposed strategy is illustrated in two real case applications using Pléiades panchromatic images: the detection of vineyards and the detection of cultivated oyster fields. In both cases, at least one of the tested multivariate models displays higher classification accuracies than gray-level cooccurrence matrix descriptors. Its high adaptability and the low number of parameters to be set are other advantages of the proposed approach.

Recent deep learning models have demonstrated strong capabilities for classifying text and non-text components in natural images. They extract a high-level feature globally computed from a whole image component (patch), where the cluttered background information may dominate true text features in the deep representation. This leads to less discriminative power and poorer robustness. In this paper, we present a new system for scene text detection by proposing a novel text-attentional convolutional neural network (Text-CNN) that particularly focuses on extracting text-related regions and features from the image components. We develop a new learning mechanism to train the Text-CNN with multi-level and rich supervised information, including text region mask, character label, and binary text/non-text information. The rich supervision information enables the Text-CNN with a strong capability for discriminating ambiguous texts, and also increases its robustness against complicated background components. The training process is formulated as a multi-task learning problem, where low-level supervised information greatly facilitates the main task of text/non-text classification. In addition, a powerful low-level detector called contrast-enhancement maximally stable extremal regions (MSERs) is developed, which extends the widely used MSERs by enhancing intensity contrast between text patterns and background. This allows it to detect highly challenging text patterns, resulting in a higher recall. Our approach achieved promising results on the ICDAR 2013 data set, with an F-measure of 0.82, substantially improving the state-of-the-art results.
The automated analysis of video captured from a first-person perspective has gained increased interest since the advent of marketed miniaturized wearable cameras. With this a person is taking visual measurements about the world in a sequence of fixations which contain relevant information about the most salient parts of the environment and the goals of the actor. We present a novel model for gaze prediction in egocentric video based on the spatiotemporal visual information captured from the wearer's camera, specifically extended using a subjective function of surprise by means of motion memory, referring to the human aspect of visual attention. Spatiotemporal saliency detection is computed in a bioinspired framework using a superposition of superpixel- and contrast based conspicuity maps as well as an optical flow based motion saliency map. Motion is further processed into a motion novelty map that is constructed by a comparison between most recent motion information with an exponentially decreasing memory of motion information. The innovative motion novelty map is experienced to be able to provide a significant increase in the performance of gaze prediction. Experimental results are gained from egocentric videos using eye-tracking glasses in a natural shopping task and prove a 6.48% increase in the mean saliency at a fixation in terms of a measure of mimicking human attention.

Detection and classification of cell nuclei in histopathology images of cancerous tissue stained with the standard hematoxylin and eosin stain is a challenging task due to cellular heterogeneity. Deep learning approaches have been shown to produce encouraging results on histopathology images in various studies. In this paper, we propose a Spatially Constrained Convolutional Neural Network (SC-CNN) to perform nucleus detection. SC-CNN regresses the likelihood of a pixel being the center of a nucleus, where high probability values are spatially constrained to locate in the vicinity of the centers of nuclei. For classification of nuclei, we propose a novel Neighboring Ensemble Predictor (NEP) coupled with CNN to more accurately predict the class label of detected cell nuclei. The proposed approaches for detection and classification do not require segmentation of nuclei. We have evaluated them on a large dataset of colorectal adenocarcinoma images, consisting of more than 20,000 annotated nuclei belonging to four different classes. Our results show that the joint detection and classification of the proposed SC-CNN and NEP produces the highest average F1 score as compared to other recently published approaches. Prospectively, the proposed methods could offer benefit to pathology practice in terms of quantitative analysis of tissue constituents in whole-slide images, and potentially lead to a better understanding of cancer.
The choice of features greatly influences the performance of a tissue classification system. Despite this, many systems are built with standard, predefined filter banks that are not optimized for that particular application. Representation learning methods such as restricted Boltzmann machines may outperform these standard filter banks because they learn a feature description directly from the training data. Like many other representation learning methods, restricted Boltzmann machines are unsupervised and are trained with a generative learning objective; this allows them to learn representations from unlabeled data, but does not necessarily produce features that are optimal for classification. In this paper we propose the convolutional classification restricted Boltzmann machine, which combines a generative and a discriminative learning objective. This allows it to learn filters that are good both for describing the training data and for classification. We present experiments with feature learning for lung texture classification and airway detection in CT images. In both applications, a combination of learning objectives outperformed purely discriminative or generative learning, increasing, for instance, the lung tissue classification accuracy by 1 to 8 percentage points. This shows that discriminative learning can help an otherwise unsupervised feature learner to learn filters that are optimized for classification.

In this paper, a novel approach for improving vehicular positioning is presented. This method is based on the cooperation of the vehicles by communicating their measured information about their position. This method consists of two steps. In the first step, we introduce our cooperative map matching method. This map matching method uses the V2V communication in a vehicular ad hoc network (VANET) to exchange global positioning system (GPS) information between vehicles. Having a precise road map, vehicles can apply the road constraints of other vehicles in their own map matching process and acquire a significant improvement in their positioning. After that, we have proposed the concept of a dynamic base station DGPS (DDGPS), which is used by vehicles in the second step to generate and broadcast the GPS pseudorange corrections that can be used by newly arrived vehicles to improve their positioning. The DDGPS is a decentralized cooperative method that aims to improve the GPS positioning by estimating and compensating the common error in GPS pseudorange measurements. It can be seen as an extension of DGPS where the base stations are not necessarily static with an exact known position. In the DDGPS method, the pseudorange corrections are estimated based on the receiver's belief on its positioning and its uncertainty and then broadcasted to other GPS receivers. The performance of the proposed algorithm has been verified with simulations in several realistic scenarios.
ETPL DIP - 026 Image De-noising Using Quadtree-Based Nonlocal Means With Locally Adaptive Principal Component Analysis

In this letter, we present an efficient image denoising method combining quadtree-based nonlocal means (NLM) and locally adaptive principal component analysis. It exploits nonlocal multiscale self-similarity better, by creating sub-patches of different sizes using quadtree decomposition on each patch. To achieve spatially uniform denoising, we propose a local noise variance estimator combined with denoiser based on locally adaptive principal component analysis. Experimental results demonstrate that our proposed method achieves very competitive denoising performance compared with state-of-the-art denoising methods, even obtaining better visual perception at high noise levels.

ETPL DIP - 027 Fusion of Multispectral and Panchromatic Images Based on Morphological Operator

Nonlinear decomposition schemes constitute an alternative to classical approaches for facing the problem of data fusion. In this paper, we discuss the application of this methodology to a popular remote sensing application called pansharpening, which consists in the fusion of a low resolution multispectral image and a high-resolution panchromatic image. We design a complete pansharpening scheme based on the use of morphological half gradient operators and demonstrate the suitability of this algorithm through the comparison with the state-of-the-art approaches. Four data sets acquired by the Pleiades, Worldview-2, Ikonos, and Geoeye-1 satellites are employed for the performance assessment, testifying the effectiveness of the proposed approach in producing top-class images with a setting independent of the specific sensor.

ETPL DIP - 028 Texture Classification Using Dense Micro-Block Difference.

This paper is devoted to the problem of texture classification. Motivated by recent advancements in the field of compressive sensing and keypoints descriptors, a set of novel features called dense micro-block difference (DMD) is proposed. These features provide highly descriptive representation of image patches by densely capturing the granularities at multiple scales and orientations. Unlike most of the earlier work on local features, the DMD does not involve any quantization, thus retaining the complete information. We demonstrate that the DMD have dimensionality much lower than Scale Invariant Feature Transform (SIFT) and can be computed using integral image much faster than SIFT. The proposed features are encoded using the Fisher vector method to obtain an image descriptor, which considers high-order statistics. The proposed image representation is combined with the linear support vector machine classifier. Extensive experiments are conducted on five texture data sets (KTH-TIPS, UMD, KTH-TIPS-2a, Brodatz, and Curet) using standard protocols. The results demonstrate that our approach outperforms the state-of-the-art in texture classification.
Among brain tumors, gliomas are the most common and aggressive, leading to a very short life expectancy in their highest grade. Thus, treatment planning is a key stage to improve the quality of life of oncological patients. Magnetic resonance imaging (MRI) is a widely used imaging technique to assess these tumors, but the large amount of data produced by MRI prevents manual segmentation in a reasonable time, limiting the use of precise quantitative measurements in the clinical practice. So, automatic and reliable segmentation methods are required; however, the large spatial and structural variability among brain tumors make automatic segmentation a challenging problem. In this paper, we propose an automatic segmentation method based on Convolutional Neural Networks (CNN), exploring small 3 x 3 kernels. The use of small kernels allows designing a deeper architecture, besides having a positive effect against overfitting, given the fewer number of weights in the network. We also investigated the use of intensity normalization as a pre-processing step, which though not common in CNN-based segmentation methods, proved together with data augmentation to be very effective for brain tumor segmentation in MRI images. Our proposal was validated in the Brain Tumor Segmentation Challenge 2013 database (BRATS 2013), obtaining simultaneously the first position for the complete, core, and enhancing regions in Dice Similarity Coefficient metric (0.88, 0.83, 0.77) for the Challenge data set. Also, it obtained the overall first position by the online evaluation platform. We also participated in the on-site BRATS 2015 Challenge using the same model, obtaining the second place, with Dice Similarity Coefficient metric of 0.78, 0.65, and 0.75 for the complete, core, and enhancing regions, respectively.

<table>
<thead>
<tr>
<th>ETPL DIP - 029</th>
<th>Texture Classification Using Dense Micro-Block Difference.</th>
</tr>
</thead>
<tbody>
<tr>
<td>This paper is devoted to the problem of texture classification. Motivated by recent advancements in the field of compressive sensing and keypoints descriptors, a set of novel features called dense micro-block difference (DMD) is proposed. These features provide highly descriptive representation of image patches by densely capturing the granularities at multiple scales and orientations. Unlike most of the earlier work on local features, the DMD does not involve any quantization, thus retaining the complete information. We demonstrate that the DMD have dimensionality much lower than Scale Invariant Feature Transform (SIFT) and can be computed using integral image much faster than SIFT. The proposed features are encoded using the Fisher vector method to obtain an image descriptor, which considers high-order statistics. The proposed image representation is combined with the linear support vector machine classifier. Extensive experiments are conducted on five texture data sets (KTH-TIPS, UMD, KTH-TIPS-2a, Brodatz, and Curet) using standard protocols. The results demonstrate that our approach outperforms the state-of-the-art in texture classification.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ETPL DIP - 030</th>
<th>Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among brain tumors, gliomas are the most common and aggressive, leading to a very short life expectancy in their highest grade. Thus, treatment planning is a key stage to improve the quality of life of oncological patients. Magnetic resonance imaging (MRI) is a widely used imaging technique to assess these tumors, but the large amount of data produced by MRI prevents manual segmentation in a reasonable time, limiting the use of precise quantitative measurements in the clinical practice. So, automatic and reliable segmentation methods are required; however, the large spatial and structural variability among brain tumors make automatic segmentation a challenging problem. In this paper, we propose an automatic segmentation method based on Convolutional Neural Networks (CNN), exploring small 3 x 3 kernels. The use of small kernels allows designing a deeper architecture, besides having a positive effect against overfitting, given the fewer number of weights in the network. We also investigated the use of intensity normalization as a pre-processing step, which though not common in CNN-based segmentation methods, proved together with data augmentation to be very effective for brain tumor segmentation in MRI images. Our proposal was validated in the Brain Tumor Segmentation Challenge 2013 database (BRATS 2013), obtaining simultaneously the first position for the complete, core, and enhancing regions in Dice Similarity Coefficient metric (0.88, 0.83, 0.77) for the Challenge data set. Also, it obtained the overall first position by the online evaluation platform. We also participated in the on-site BRATS 2015 Challenge using the same model, obtaining the second place, with Dice Similarity Coefficient metric of 0.78, 0.65, and 0.75 for the complete, core, and enhancing regions, respectively.</td>
<td></td>
</tr>
</tbody>
</table>
**Distance metric learning (DML) is an important technique to improve similarity search in content-based image retrieval. Despite being studied extensively, most existing DML approaches typically adopt a single-modal learning framework that learns the distance metric on either a single feature type or a combined feature space where multiple types of features are simply concatenated. Such single-modal DML methods suffer from some critical limitations: (i) some type of features may significantly dominate the others in the DML task due to diverse feature representations; and (ii) learning a distance metric on the combined high-dimensional feature space can be extremely time-consuming using the naive feature concatenation approach. To address these limitations, in this paper, we investigate a novel scheme of online multi-modal distance metric learning (OMDML), which explores a unified two-level online learning scheme: (i) it learns to optimize a distance metric on each individual feature space; and (ii) then it learns to find the optimal combination of diverse types of features. To further reduce the expensive cost of DML on high-dimensional feature space, we propose a low-rank OMDML algorithm which not only significantly reduces the computational cost but also retains highly competing or even better learning accuracy. We conduct extensive experiments to evaluate the performance of the proposed algorithms for multi-modal image retrieval, in which encouraging results validate the effectiveness of the proposed technique.**

**Stereoscopic image retargeting plays an important role in adaptive 3D stereoscopic displays. It aims to fit displays with various resolutions while preserving visually salient content and geometric consistency. We propose a stereoscopic image retargeting method based on stereoscopic visual attention guided seam carving. Firstly, stereoscopic saliency map is generated by combining 2D saliency and depth saliency maps, and significant energy map is generated by considering binocular disparity binocular and binocular just-noticeable-difference (BJND). Then, seam selection is applied to the left image based on stereoscopic saliency and energy maps, and seam replacement is performed for the occluded regions to prevent the geometry inconsistency. Finally, according to the matched left and right seams, the retargeted stereoscopic image is generated. In the experiments, subjective and objective analysis on three stereoscopic image databases shows that the proposed approach produces better seam carving results than the related existing methods.**
The preservation of image quality under various display conditions becomes more and more important in the multimedia era. A considerable amount of effort has been devoted to compensating the quality degradation caused by dim LCD backlight for mobile devices and desktop monitors. However, most previous enhancement methods for backlight-scaled images only consider the luminance component and overlook the impact of color appearance on image quality. In this paper, we propose a fast and elegant method that exploits the anchoring property of human visual system to preserve the color appearance of backlight-scaled images as much as possible. Our approach is distinguished from previous ones in many aspects. First, it has a sound theoretical basis. Second, it takes the luminance and chrominance components into account in an integral manner. Third, it has low complexity and can process 720p high-definition videos at 35 frames per second without flicker. The superior performance of the proposed method is verified through psychophysical tests.

Least Significant Bit (LSB) steganography is a well-known technique which operates in the spatial domain of digital images. In this paper, the LSB Word-Hunt (LSB WH) is presented. It is a novel LSB approach inspired by the word-hunt puzzle. The main focus of LSB WH is to reduce the Expected Number of Modifications per Pixel (ENMPP) when compared to other methods in the literature. The results show that LSB WH has an ENMPP around 0.315, for natural images with high entropy on the second and the third least significant bits. Results also show that the new method is robust to the statistical chi-square attack.

Magnetic resonance (MR) imaging is vulnerable to a variety of artifacts, which potentially degrade the perceived quality of MR images and, consequently, may cause inefficient and/or inaccurate diagnosis. In general, these artifacts can be classified as structured or unstructured depending on the correlation of the artifact with the original content. In addition, the artifact can be white or colored depending on the flatness of the frequency spectrum of the artifact. In current MR imaging applications, design choices allow one type of artifact to be traded off with another type of artifact. Hence, to support these design choices, the relative impact of structured versus unstructured or colored versus white artifacts on perceived image quality needs to be known. To this end, we conducted two subjective experiments. Clinical application specialists rated the quality of MR images, distorted with different types of artifacts at various levels of degradation. The results demonstrate that unstructured artifacts deteriorate quality less than structured artifacts, while colored artifacts preserve quality better than white artifacts.
Automatic segmentation of the primary object in a video clip is a challenging problem as there is no prior knowledge of the primary object. Most existing techniques thus adapt an iterative approach for foreground and background appearance modeling, i.e., fix the appearance model while optimizing the segmentation and fix the segmentation while optimizing the appearance model. However, these approaches may rely on good initialization and can be easily trapped in local optimal. In addition, they are usually time consuming for analyzing videos. To address these limitations, we propose a novel and efficient appearance modeling technique for automatic primary video object segmentation in the Markov random field (MRF) framework. It embeds the appearance constraint as auxiliary nodes and edges in the MRF structure, and can optimize both the segmentation and appearance model parameters simultaneously in one graph cut. The extensive experimental evaluations validate the superiority of the proposed approach over the state-of-the-art methods, in both efficiency and effectiveness.

In this paper, we introduce a novel approach for active contours with free endpoints. A scheme for image segmentation is presented based on a discrete version of the Mumford-Shah functional where the contours can be both closed and open curves. Additional to a flow of the curves in normal direction, evolution laws for the tangential flow of the endpoints are derived. Using a parametric approach to describe the evolving contours together with an edge-preserving denoising, we obtain a fast method for image segmentation and restoration. The analytical and numerical schemes are presented followed by numerical experiments with artificial test images and with a real medical image.

A novel biometric face recognition algorithm using depth cameras is proposed. The key contribution is the design of a novel and highly discriminative face image descriptor called bag of dense derivative depth patterns (Bag-D3P). This descriptor is composed of four different stages that fully exploit the characteristics of depth information: 1) dense spatial derivatives to encode the 3-D local structure; 2) face-adaptive quantization of the previous derivatives; 3) multibag of words that creates a compact vector description from the quantized derivatives; and 4) spatial block division to add global spatial information. The proposed system can recognize people faces from a wide range of poses, not only frontal ones, increasing its applicability to real situations. Last, a new face database of high-resolution depth images has been created and made it public for evaluation purposes.
In multimedia information retrieval, most classic approaches tend to represent different modalities of media in the same feature space. With the click data collected from the users' searching behavior, existing approaches take either one-to-one paired data (text-image pairs) or ranking examples (text-query-image and/or image-query-text ranking lists) as training examples, which do not make full use of the click data, particularly the implicit connections among the data objects. In this paper, we treat the click data as a large click graph, in which vertices are images/text queries and edges indicate the clicks between an image and a query. We consider learning a multimodal representation from the perspective of encoding the explicit/implicit relevance relationship between the vertices in the click graph. By minimizing both the truncated random walk loss as well as the distance between the learned representation of vertices and their corresponding deep neural network output, the proposed model which is named multimodal random walk neural network (MRW-NN) can be applied to not only learn robust representation of the existing multimodal data in the click graph, but also deal with the unseen queries and images to support cross-modal retrieval. We evaluate the latent representation learned by MRW-NN on a public large-scale click log data set Clickture and further show that MRW-NN achieves much better cross-modal retrieval performance on the unseen queries/images than the other state-of-the-art methods.

Semantic image segmentation is a fundamental yet challenging problem, which can be viewed as an extension of the conventional object detection with close relation to image segmentation and classification. It aims to partition images into non-overlapping regions that are assigned predefined semantic labels. Most of the existing approaches utilize and integrate low-level local features and high-level contextual cues, which are fed into an inference framework such as, the conditional random field (CRF). However, the lack of meaning in the primitives (i.e., pixels or superpixels) and the cues provides low discriminatory capabilities, since they are rarely object-consistent. Moreover, blind combinations of heterogeneous features and contextual cues exploitation through limited neighborhood relations in the CRFs tend to degrade the labeling performance. This paper proposes an ontology-based semantic image segmentation (OBSIS) approach that jointly models image segmentation and object detection. In particular, a Dirichlet process mixture model transforms the low-level visual space into an intermediate semantic space, which drastically reduces the feature dimensionality. These features are then individually weighed and independently learned within the context, using multiple CRFs. The segmentation of images into object parts is hence reduced to a classification task, where object inference is passed to an ontology model. This model resembles the way by which humans understand the images through the combination of different cues, context models, and rule-based learning of the ontologies. Experimental evaluations using the MSRC-21 and PASCAL VOC’2010 data sets show promising results.
## ETPL DIP - 041  Robust Texture Image Representation by Scale Selective Local Binary Patterns.

Local binary pattern (LBP) has successfully been used in computer vision and pattern recognition applications, such as texture recognition. It could effectively address grayscale and rotation variation. However, it failed to get desirable performance for texture classification with scale transformation. In this paper, a new method based on dominant LBP in scale space is proposed to address scale variation for texture classification. First, a scale space of a texture image is derived by a Gaussian filter. Then, a histogram of pre-learned dominant LBPs is built for each image in the scale space. Finally, for each pattern, the maximal frequency among different scales is considered as the scale invariant feature. Extensive experiments on five public texture databases (University of Illinois at Urbana-Champaign, Columbia Utrecht Database, Kungliga Tekniska Högskolan-Textures under varying Illumination, Pose and Scale, University of Maryland, and Amsterdam Library of Textures) validate the efficiency of the proposed feature extraction scheme. Coupled with the nearest subspace classifier, the proposed method could yield competitive results, which are 99.36%, 99.51%, 99.39%, 99.46%, and 99.71% for UIUC, CUReT, KTH-TIPS, UMD, and ALOT, respectively. Meanwhile, the proposed method inherits simple and efficient merits of LBP, for example, it could extract scale-robust feature for a $200 \times 200$ image within 0.24 s, which is applicable for many real-time applications.

## ETPL DIP - 042  Predicting Vascular Plant Richness in a Heterogeneous Wetland Using Spectral and Textural Features and a Random Forest Algorithm.

A method to predict vascular plant richness using spectral and textural variables in a heterogeneous wetland is presented. Plant richness was measured at 44 sampling plots in a 16-ha anthropogenic peatland. Several spectral indices, first-order statistics (median and standard deviation), and second-order statistics [metrics of a gray-level co-occurrence matrix (GLCM)] were extracted from a Landsat 8 Operational Land Imager image and a Pleiades 1B image. We selected the most important variables for predicting richness using recursive feature elimination and then built a model using random forest regression. The final model was based on only two textural variables obtained from the GLCM and derived from the Landsat 8 image. An accurate predictive capability was reported ($R^2 = 0.6$; RMSE = 1.99 species), highlighting the possibility of obtaining parsimionous models using textural variables. In addition, the results showed that the mid-resolution Landsat 8 image provided better predictors of richness than the high-resolution Pleiades image. This is the first study to generate a model for plant richness in a wetland ecosystem.
The local variance of image intensity is a typical measure of image smoothness. It has been extensively used, for example, to measure the visual saliency or to adjust the filtering strength in image processing and analysis. However, to the best of our knowledge, no analytical work has been reported about the effect of JPEG compression on image local variance. In this paper, a theoretical analysis on the variation of local variance caused by JPEG compression is presented. First, the expectation of intensity variance of 8×8 non-overlapping blocks in a JPEG image is derived. The expectation is determined by the Laplacian parameters of the discrete cosine transform coefficient distributions of the original image and the quantization step sizes used in the JPEG compression. Second, some interesting properties that describe the behavior of the local variance under different degrees of JPEG compression are discussed. Finally, both the simulation and the experiments are performed to verify our derivation and discussion. The theoretical analysis presented in this paper provides some new insights into the behavior of local variance under JPEG compression. Moreover, it has the potential to be used in some areas of image processing and analysis, such as image enhancement, image quality assessment, and image filtering.

This paper presents a novel rank-based method for image watermarking. In the watermark embedding process, the host image is divided into blocks, followed by the 2-D discrete cosine transform (DCT). For each image block, a secret key is employed to randomly select a set of DCT coefficients suitable for watermark embedding. Watermark bits are inserted into an image block by modifying the set of DCT coefficients using a rank-based embedding rule. In the watermark detection process, the corresponding detection matrices are formed from the received image using the secret key. Afterward, the watermark bits are extracted by checking the ranks of the detection matrices. Since the proposed watermarking method only uses two DCT coefficients to hide one watermark bit, it can achieve very high embedding capacity. Moreover, our method is free of host signal interference. This desired feature and the usage of an error buffer in watermark embedding result in high robustness against attacks. Theoretical analysis and experimental results demonstrate the effectiveness of the proposed method.
In this paper, we are concerned with the problem of automatic scene text recognition, which involves localizing and reading characters in natural images. We investigate this problem from the perspective of representation and propose a novel multi-scale representation, which leads to accurate, robust character identification and recognition. This representation consists of a set of mid-level primitives, termed strokelets, which capture the underlying substructures of characters at different granularities. The Strokelets possess four distinctive advantages: 1) usability: automatically learned from character level annotations; 2) robustness: insensitive to interference factors; 3) generality: applicable to variant languages; and 4) expressivity: effective at describing characters. Extensive experiments on standard benchmarks verify the advantages of the strokelets and demonstrate the effectiveness of the text recognition algorithm built upon the strokelets. Moreover, we show the method to incorporate the strokelets to improve the performance of scene text detection.

In this paper, we consider an image decomposition model that provides a novel framework for image denoising. The model computes the components of the image to be processed in a moving frame that encodes its local geometry (directions of gradients and level lines). Then, the strategy we develop is to denoise the components of the image in the moving frame in order to preserve its local geometry, which would have been more affected if processing the image directly. Experiments on a whole image database tested with several denoising methods show that this framework can provide better results than denoising the image directly, both in terms of Peak signal-to-noise ratio and Structural similarity index metrics.
Feature learning is an intensively studied research topic in image classification. Although existing methods like sparse coding, locality-constrained linear coding, fisher vector encoding, etc., have shown their effectiveness in image representation, most of them overlook a phenomenon called the small sample size problem, where the number of training samples is relatively smaller than the dimensionality of the features, which may limit the predictive power of the classifier. Subspace learning is a strategy to mitigate this problem by reducing the dimensionality of the features. However, most conventional subspace learning methods attempt to learn a global subspace to discriminate all the classes, which proves to be difficult and ineffective in multi-class classification task. To this end, we propose to learn a local subspace for each sample instead of learning a global subspace for all samples. Our key observation is that, in multi-class image classification, the label of each testing sample is only confused by a few classes which have very similar visual appearance to it. Thus, in this work, we propose a coarse-to-fine strategy, which first picks out such classes, and then conducts a local subspace learning to discriminate them. As the subspace learning method is regularized and conducted within some selected classes, we term it selective regularized subspace learning (SRSL), and we term our classification pipeline selective regularized subspace learning based multi-class image classification (SRSL_MIC). Experimental results on four representative datasets (Caltech-101, Indoor-67, ORL Faces and AR Faces) demonstrate the effectiveness of the proposed method.
Thank You!