



# ELYSIUM PRO

A UNIT OF ELYSIUM GROUPS

## FINAL YEAR PROJECT

BIO-INFORMATICS 2019-2020

TITLES WITH ABSTRACTS



CALL US @

(+91) 9944 7933 98 | (+91) 452 - 424 2842, 424 2843

20 Years of Experience | Automated Services | 24/7 Help Desk Support  
Advanced Technologies and Tools | Legitimate Members of all Journals  
Quality Project Training | Industry Exposure

Elysium PRO

Titles with Abstracts 2019-20





**ELYSIUMPRO**  
INSPIRING THE LEADING EDGE TECHNOLOGIES

[www.elysiumpro.in](http://www.elysiumpro.in)



#227, Elysium Campus, Church Road, Anna Nagar,  
Madurai - 625020, Tamil Nadu, India

CALL US @

**(+91) 9944 7933 98 | (+91) 452 - 424 2842, 424 2843**

[f /ElysiumPro Project Center](#)

[t /ElysiumPro](#)

[in /ElysiumPro](#)

Elysium PRO

Titles with Abstracts 2019-20



**EPRO BIO-  
INFO - 001**

**Efficient Face Recognition Using Regularized Adaptive Non-Local Sparse Coding**

In the sparse representation-based classification (SRC), the object recognition procedure depends on local sparsity identification from sparse coding coefficients, where many existing SRC methods have focused on the local sparsity and the samples correlation to improve the classifier performance. However, the coefficients often do not accurately represent the local sparsity due to several factors that affect the data acquisition process such as noise, blurring, and downsampling. Therefore, this paper presents an effective method that exploits nonlocal sparsity by estimating the sparse code changes, which can be done by adding a nonlocal constraint term to the local constraint one. In addition, for generality, the sparse coding and regularization parameters are adaptively estimated. A comparative study demonstrated that the proposed method has better accuracy rates compared to the existing state-of-the-art methods.

**EPRO BIO-  
INFO - 002**

**Cascaded Static and Dynamic Local Feature Extractions for Face Sketch to Photo Matching**

The automatic identification of a corresponding photo from a face sketch can assist in criminal investigations. The face sketch is rendered based on the descriptions elicited by the eyewitness. This may cause the face sketch to have some degrees of shape exaggeration that make some parts of the face geometrically misaligned. In this paper, we attempt to address the effect of these influences by a cascaded static and dynamic local feature extraction method so that the constructed feature vectors are built based on the correct patches. In the proposed method, the feature vectors from the local static extraction on a sketch and photo are matched using the nearest neighbors. Then, some  $n$  most similar photos are shortlisted based on the nearest neighbors. These photos are eventually re-matched using feature vectors from the local dynamic extraction method. The feature vectors are matched using the  $L_1$ -distance measure. The experimental results for The Chinese University of Hong Kong (CUHK) Face Sketch Database (CUFS) and CUHK Face Sketch FERET Database (CUFSF) datasets indicate that the proposed method outperforms the state-of-the-art methods.

**EPRO BIO-  
INFO - 003**

**Generating Photographic Faces from the Sketch Guided by Attribute Using GAN**

From a sketch image or text description, generating a semantic and photographic face image has always been an extremely important issue in computer vision. Sketch images generally contain only simple profile information but not the detail of the face. Therefore, it is difficult to generate facial attributes accurately. In this paper, we treat the sketch to face the problem as a face hallucination reconstruction problem. In order to solve this problem, we propose an image translation network by exploiting attributes with the generated adversarial network. And it can significantly contribute to the authenticity of the generated face by supplementing sketch image with the additional facial attribute feature. The generator network is composed of a feature extracting network and downsampling-upsampling network, both networks use skip-connection to reduce the number of layers without affecting network performance. The discriminator network is designed to examine whether the generated faces contain the desired attributes or not. In the underlying feature extraction phase, our network is different from most attribute-embedded networks, we fuse the sketch images and attributes perceptually. We set the network sub-Branch A and B, which receive a sketch image and attribute vector in order to extract low-level profile information and high-level semantic features. Compared with the state-of-the-art methods of image translation, the performance of the proposed network is excellent.

**EPRO BIO-  
INFO - 004**

**Dictionary Representation of Deep Features for Occlusion-Robust Face Recognition**

Deep learning has achieved exciting results in face recognition; however, the accuracy is still unsatisfying for occluded faces. To improve the robustness for occluded faces, this paper proposes a novel deep dictionary representation-based classification scheme, where a convolutional neural network is employed as the feature extractor and followed by a dictionary to linearly code the extracted deep features. The dictionary is composed by a gallery part consisting of the deep features of the training samples and an auxiliary part consisting of the mapping vectors acquired from the subjects either inside or outside the training set and associated with the occlusion patterns of the testing face samples. A squared Euclidean norm is used to regularize the coding coefficients. The proposed scheme is computationally efficient and is robust to large contiguous occlusion. In addition, the proposed scheme is generic for both the occluded and non-occluded face images and works with a single training sample per subject. The extensive experimental evaluations demonstrate the superior performance of the proposed approach over other state-of-the-art algorithms.

**EPRO BIO-  
INFO - 005**

**Graph Fusion for Finger Multimodal Biometrics**

In terms of biometrics, a human finger itself is with trimodal traits including fingerprint, finger-vein, and finger-knuckle-print, which provides convenience and practicality for finger trimodal fusion recognition. The scale inconsistency of finger trimodal images is an important reason affecting effective fusion. It is therefore very important to developing a theory of giving a unified expression of finger trimodal features. In this paper, a graph-based feature extraction method for finger biometric images is proposed. The feature expression based on graph structure can well solve the problem of feature space mismatch for the finger three modalities. We provide two fusion frameworks to integrate the finger trimodal graph features together, the serial fusion and coding fusion. The research results can not only promote the advancement of finger multimodal biometrics technology but also provide a scientific solution framework for other multimodal feature fusion problems. The experimental results show that the proposed graph fusion recognition approach obtains a better and more effective recognition performance in finger biometrics.

**EPRO BIO-  
INFO - 006**

**Learning Affective Video Features for Facial Expression Recognition via Hybrid Deep Learning**

One key challenging issues of facial expression recognition (FER) in video sequences is to extract discriminative spatiotemporal video features from facial expression images in video sequences. In this paper, we propose a new method of FER in video sequences via a hybrid deep learning model. The proposed method first employs two individual deep convolutional neural networks (CNNs), including a spatial CNN processing static facial images and a temporal CN network processing optical flow images, to separately learn high-level spatial and temporal features on the divided video segments. These two CNNs are fine-tuned on target video facial expression datasets from a pre-trained CNN model. Then, the obtained segment-level spatial and temporal features are integrated into a deep fusion network built with a deep belief network (DBN) model. This deep fusion network is used to jointly learn discriminative spatiotemporal features. Finally, an average pooling is performed on the learned DBN segment-level features in a video sequence, to produce a fixed-length global video feature representation. Based on the global video feature representations, a linear support vector machine (SVM) is employed for facial expression classification tasks. The extensive experiments on three public video-based facial expression datasets, i.e., BAUM-1s, RML, and MMI, show the effectiveness of our proposed method, outperforming the state-of-the-arts.

**EPRO BIO-  
INFO - 007**

**Learning Low-Rank Regularized Generic Representation with Block-Sparse Structure for Single Sample Face Recognition**

In this paper, we propose a novel low-rank regularized generic representation method to address the single sample per person problem in face recognition, which simultaneously employs the structure information from the probe dataset and the generic variation information. Each face image is divided into overlapped patches, whose classification results are aggregated to produce the final result. We generate a subspace for each patch by extracting its eight nearest neighbor patches and explore the relationships between subspaces by imposing a low-rank constraint on the reconstruction coefficients. Moreover, a block-sparsity constraint on the coefficient matrix is imposed to further promote the discrimination of representations. We also propose a dictionary learning method to learn the intra-class facial variations from the generic face datasets, which separates the face contour noise from the variation dictionary by an incoherence regularization item. The experimental results on four public face databases not only show the robustness of our approach to expression, illumination, occlusion, and time variation but also demonstrate the effectiveness of face sketch recognition.

**EPRO BIO-  
INFO - 008**

**Short-Range Radar Based Real-Time Hand Gesture Recognition Using LSTM Encoder**

Due to the development of short-range radar with high-resolution, the radar sensor has a high potential to be used in real human-computer interaction (HCI) applications. The radar sensor has advantages over optical cameras in that it is unaffected by illumination and it is able to detect the objects in an occluded environment. This paper proposes a hand gesture recognition system for a real-time application of HCI using 60 GHz frequency-modulated continuous wave (FMCW) radar, Soli, developed by Google. The overall system includes signal processing part that generates range-Doppler map (RDM) sequences without clutter and machine learning part including a long short-term memory (LSTM) encoder to learn the temporal characteristics of the RDM sequences. A set of data is collected from 10 participants for the experiment. The proposed hand gesture recognition system successfully distinguishes 10 gestures with a high classification accuracy of 99.10%. It also recognizes the gestures of a new participant with an accuracy of 98.48%.

**EPRO BIO-  
INFO - 009**

**Dynamic Sign Language Recognition Based on Video Sequence with BLSTM-3D Residual Networks**

Sign language recognition aims to recognize meaningful movements of hand gestures and is a significant solution in intelligent communication between the deaf community and hearing societies. However, until now, the current dynamic sign language recognition methods still have some drawbacks with difficulties of recognizing complex hand gestures, low recognition accuracy for most dynamic sign language recognition, and potential problems in larger video sequence data training. In order to solve these issues, this paper presents a multimodal dynamic sign language recognition method based on a deep 3-dimensional residual ConvNet and bi-directional LSTM networks, which is named as BLSTM-3D residual network (B3D ResNet). This method consists of three main parts. First, the hand object is localized in the video frames in order to reduce the time and space complexity of network calculation. Then, the B3D ResNet automatically extracts the spatiotemporal features from the video sequences and establishes an intermediate score corresponding to each action in the video sequence after feature analysis. Finally, by classifying the video sequences, the dynamic sign language is accurately identified. The experiment is conducted on test datasets, including DEVISIGN\_D dataset and SLR\_Dataset. The results show that the proposed method can obtain state-of-the-art recognition accuracy (89.8% on the DEVISIGN\_D dataset and 86.9% on SLR\_Dataset).

**EPRO BIO-  
INFO - 010**

**Efficient Facial Expression Recognition Algorithm Based on Hierarchical Deep Neural Network Structure**

With the continued development of artificial intelligence (AI) technology, research on interaction technology has become more popular. Facial expression recognition (FER) is an important type of visual information that can be used to understand a human's emotional situation. In particular, the importance of AI systems has recently increased due to advancements in research on AI systems applied to AI robots. In this paper, we propose a new scheme for FER system based on hierarchical deep learning. The feature extracted from the appearance feature-based network is fused with the geometric feature in a hierarchical structure. The appearance feature-based network extracts holistic features of the face using the preprocessed LBP image, whereas the geometric feature-based network learns the coordinate change of action units (AUs) landmark, which is a muscle that moves mainly when making facial expressions. The proposed method combines the result of the softmax function of two features by considering the error associated with the second highest emotion (Top-2) prediction result. In addition, we propose a technique to generate facial images with neutral emotion using the autoencoder technique. By this technique, we can extract the dynamic facial features between the neutral and emotional images without sequence data.

**EPRO BIO-  
INFO - 011**

**Single-Sample Face Recognition Based on Feature Expansion**

Face recognition (FR) with a single sample per person (SSPP) is one of the most challenging problems in computer vision. In this scenario, it is difficult to predict facial variation such as pose, illumination, and disguise due to the lack of enough training samples. Therefore, the development of the FR system with only a small number of training samples is hindered. To address this problem, this paper proposes a scheme combined transfer learning and sample expansion in feature space. First, it uses transfer learning by training a deep convolutional neural network on a common multi-sample face dataset and then applies the well-trained model to a target data set. Second, it proposes a sample expansion method in feature space called k class feature transfer (KCFT) to enrich intra-class variation information for a single-sample face feature. Compared with other expanding sample methods in the image domain, this method of expanding the samples in the feature domain is novel and easy to implement. Third, it trains a softmax classifier with expanded face features. The experimental results on ORL, FERET, and LFW face databases demonstrate the effectiveness and robustness of the proposed method for various facial variations.

**EPRO BIO-  
INFO – 012**

**Head Pose Estimation in the Wild Assisted by Facial Landmarks Based on Convolutional Neural Networks**

Convolutional neural networks (CNNs) exhibit excellent performance on the head pose estimation problem under controllable conditions, but their generalization ability in the wild needs to be improved. To address this issue, we propose an approach involving the introduction of facial landmark information into the task simplifier and landmark heatmap generator constructed before the feed-forward neural network, which can use this information to normalize the face shape into a canonical shape and generate a landmark heatmap based on the transformed facial landmarks to assist in feature extraction, for enhancing generalization ability in the wild. Our method was trained on 300W-LP and tested on AFLW2000-3D. The result shows that for the same feed-forward neural network when our method is used to introduce facial landmark information into a CNN, accuracy improves from 88.5% to 99.0% and mean average error decreases from 5.94° to 1.46° on AFLW2000-3D. Furthermore, we evaluate our method on several datasets used for pose estimation and compare the result with AFLW2000-3D, finding that the features extracted by a CNN could not reflect the head pose efficiently, which limits the performance of the CNN on the head pose estimation problem in wild. By introducing facial landmarks, the CNN could extract features that reflect head pose more efficiently, thereby significantly improving the accuracy of head pose estimation in the wild.

**EPRO BIO-  
INFO - 013**

**Self-Residual Attention Network for Deep Face Recognition**

Discriminative feature embedding is of essential importance in the field of large scale face recognition. In this paper, we propose a self residual attention-based convolutional neural network (SRANet) for discriminative face feature embedding, which aims to learn the long-range dependencies of face images by decreasing the information redundancy among channels and focusing on the most informative components of spatial feature maps. More specifically, the proposed attention module consists of the self channel attention (SCA) block and self spatial attention (SSA) block which adaptively aggregates the feature maps in both channel and spatial domains to learn the inter-channel relationship matrix and the inter-spatial relationship matrix; moreover, matrix multiplications are conducted for a refined and robust face feature. With the attention module we proposed, we can make standard convolutional neural networks (CNNs), such as ResNet-50 and ResNet-101, which have more discriminative power for deep face recognition. The experiments on Labelled Faces in the Wild (LFW), Age Database (AgeDB), Celebrities in Frontal Profile (CFP), and MegaFace Challenge 1 (MF1) show that our proposed SRANet structure consistently outperforms naive CNNs and achieves state-of-the-art performance.

**EPRO BIO-  
INFO - 014**

**Local Learning with Deep and Handcrafted Features for Facial Expression Recognition**

We present an approach that combines automatic features learned by convolutional neural networks (CNN) and handcrafted features computed by the bag-of-visual-words (BOVW) model in order to achieve the state-of-the-art results in facial expression recognition (FER). To obtain automatic features, we experiment with multiple CNN architectures, pre-trained models, and training procedures, e.g., Dense-Sparse-Dense. After fusing the two types of features, we employ a local learning framework to predict the class label for each test image. The local learning framework is based on three steps. First, a k-nearest neighbors model is applied in order to select the nearest training samples for an input test image. Second, a one-versus-all support vector machines (SVM) classifier is trained on the selected training samples. Finally, the SVM classifier is used to predict the class label only for the test image it was trained for. Although we have used local learning in combination with handcrafted features in our previous work, to the best of our knowledge, local learning has never been employed in combination with deep features. The experiments on the 2013 FER Challenge data set, the FER+ data set, and the AffectNet data set demonstrate that our approach achieves the state-of-the-art results. With a top accuracy of 75.42% on the FER 2013, 87.76% on the FER+, 59.58% on the AffectNet eight-way classification, and 63.31% on the AffectNet seven-way classification, we surpass the state-of-the-art methods by more than 1% on all data sets.

**EPRO BIO-  
INFO - 015**

**Classroom Micro-Expression Recognition Algorithms Based on Multi-Feature Fusion**

The extraction of facial features is the key to micro-expression recognition. This paper puts forward a micro-expression recognition algorithm through multi-feature fusion. In this algorithm, the change of local binary pattern (LBP) feature distribution is correlated with projection error. For fast and accurate detection, the research data were all extracted from professional facial expression databases, the images had basically the same positions in each expression library, and the pure face image was acquired through manual segmentation from the selected expression library. Through comparison and its application in an intelligent classroom environment, it is proved that the proposed algorithm clearly outperforms the original LBP algorithm. The proposed method can be extended to other image recognition and classification problems.

**EPRO BIO-  
INFO - 016**

**Deep Unified Model for Face Recognition Based on Convolution Neural Network and Edge Computing**

Currently, data generated by smart devices connected through the Internet is increasing relentlessly. An effective and efficient paradigm is needed to deal with the bulk amount of data produced by the Internet of Things (IoT). Deep learning and edge computing are the emerging technologies, which are used for efficient processing of huge amount of data with distinct accuracy. In this world of advanced information systems, one of the major issues is authentication. Several techniques have been employed to solve this problem. Face recognition is considered as one of the most reliable solutions. Usually, for face recognition, scale-invariant feature transforms (SIFT) and speeded up robust features (SURF) have been used by the research community. This paper proposes an algorithm for face detection and recognition based on convolution neural networks (CNN), which outperform the traditional techniques. In order to validate the efficiency of the proposed algorithm, a smart classroom for the student's attendance using face recognition has been proposed. The face recognition system is trained on publically available labeled faces in the wild (LFW) dataset. The system can detect approximately 35 faces and recognizes 30 out of them from the single image of 40 students. The proposed system achieved 97.9% accuracy on the testing data. Moreover, generated data by smart classrooms is computed and transmitted through an IoT-based architecture using edge computing. A comparative performance study shows that our architecture outperforms in terms of data latency and real-time response.

**EPRO BIO-  
INFO - 017**

**Deep Residual Network-Based Recognition of Finger Wrinkles Using Smartphone Camera**

Iris, fingerprint, and three-dimensional face recognition technologies used in mobile devices face obstacles owing to price and size restrictions by additional cameras, lighting, and sensors. As an alternative, two-dimensional face recognition based on the built-in visible-light camera of mobile devices has been widely used. However, face recognition performance is greatly influenced by the factors, such as facial expression, illumination, and pose changes. Considering these limitations, researchers have studied palmprint, touchless fingerprint, and finger-knuckle-print recognition using the built-in visible light camera. However, these techniques reduce user convenience because of the difficulty in positioning a palm or fingers on the camera. To consider these issues, we propose a biometric system based on a finger-wrinkle image acquired by the visible-light camera of a smartphone. A deep residual network is used to address the degradation of recognition performance caused by misalignment and illumination variation occurring during image acquisition. Owing to the unavailability of the finger-wrinkle open database obtained by smartphone camera, we built the Dongguk finger-wrinkle database, including the images from 33 people. The results show that the recognition performance by our method exceeds in those of conventional methods.

**EPRO BIO-  
INFO - 018**

**Uniform and Variational Deep Learning for RGB-D Object Recognition and Person Re-Identification**

In this paper, we propose a uniform and variational deep learning (UVDL) method for RGB-D object recognition and person re-identification. Unlike most existing object recognition and person re-identification methods which usually only use the visual appearance information from RGB images, our method recognizes visual objects and persons with RGB-D images to exploit more reliable information such as geometric and anthropometric information which are robust to different viewpoints. Specifically, we extract the depth feature and the appearance feature from the depth and RGB images with two deep convolutional neural networks, respectively. In order to combine the depth feature and the appearance feature to exploit their relationship, we design a uniform and variational multimodal auto-encoder at the top layer of our deep network to seek a uniform latent variable by projecting them into a common space, which contains the whole information of RGB-D images and has small intra-class variation and large inter-class variation, simultaneously. Finally, we optimize the auto-encoder layer and two deep convolutional neural networks jointly to minimize the discriminative loss and the reconstruction error. Experimental results on both RGB-D object recognition and RGB-D person reidentification are presented to show the efficiency of our proposed approach.

**EPRO BIO-  
INFO - 019**

**Person Recognition in Personal Photo Collections**

People nowadays share large parts of their personal lives through social media. Being able to automatically recognise people in personal photos may greatly enhance user convenience by easing photo album organisation. For human identification task, however, traditional focus of computer vision has been face recognition and pedestrian re-identification. Person recognition in social media photos sets new challenges for computer vision, including non-cooperative subjects (e.g. backward viewpoints, unusual poses) and great changes in appearance. To tackle this problem, we build a simple person recognition framework that leverages convnet features from multiple image regions (head, body, etc.). We propose new recognition scenarios that focus on the time and appearance gap between training and testing samples. We present an in-depth analysis of the importance of different features according to time and viewpoint generalisability. In the process, we verify that our simple approach achieves the state of the art result on the PIPA [1] benchmark, arguably the largest social media based benchmark for person recognition to date with diverse poses, viewpoints, social groups, and events.

**EPRO BIO-  
INFO – 020**

**Context Based Emotion Recognition using EMOTIC Dataset**

In our everyday lives and social interactions we often try to perceive the emotional states of people. There has been a lot of research in providing machines with a similar capacity of recognizing emotions. From a computer vision perspective, most of the previous efforts have been focusing in analyzing the facial expressions and, in some cases, also the body pose. Some of these methods work remarkably well in specific settings. However, their performance is limited in natural, unconstrained environments. Psychological studies show that the scene context, in addition to facial expression and body pose, contributes important information to our perception of people's emotions. However, the processing of the context for automatic emotion recognition has not been explored in depth, partly due to the lack of proper data. In this paper we present EMOTIC, a dataset of images of people in natural and different situations annotated with their apparent emotion. The EMOTIC database combines two different types of emotion representation: (1) a set of 26 discrete categories, and (2) the continuous dimensions Valence, Arousal, and Dominance. We also present a detailed statistical and algorithmic analysis of the dataset along with annotators' agreement analysis. Using the EMOTIC database we train different CNN models for emotion recognition, combining the information of the person bounding box with the information present in the scene context. Our results show how scene context contributes important information to automatically recognize emotional states and motivate further research in this direction.

**EPRO BIO-  
INFO - 021**

**Discriminant Functional Learning of Color Features for the Recognition of Facial Action Units and their Intensities**

Color is a fundamental image feature of facial expressions. For example, when we furrow our eyebrows in anger, blood rushes in and a reddish color becomes apparent around that area of the face. Surprisingly, these image properties have not been exploited to recognize the facial action units (AUs) associated with these expressions. Herein, we present the first system to do recognition of AUs and their intensities using these functional color changes. These color features are shown to be robust to changes in identity, gender, race, ethnicity and skin color. Specifically, we identify the chromaticity changes defining the transition of an AU from inactive to active and use an innovative Gabor transform-based algorithm to gain invariance to the timing of these changes. Because these image changes are given by functions rather than vectors, we use a functional classifiers to identify the most discriminant color features of an AU and its intensities. We demonstrate that, using these discriminant color features, one can achieve results superior to those of the state-of-the-art. Finally, we define an algorithm that allows us to use the learned functional color representation in still images. This is done by learning the mapping between images and the identified functional color features in videos. Our algorithm works in realtime, i.e., >30 frames/second/CPU thread

**EPRO BIO-  
INFO - 022**

**Representation Learning by Rotating Your Faces**

The large pose discrepancy between two face images is one of the fundamental challenges in automatic face recognition. Conventional approaches to pose-invariant face recognition either perform face frontalization on, or learn a pose-invariant representation from, a non-frontal face image. We argue that it is more desirable to perform both tasks jointly to allow them to leverage each other. To this end, this paper proposes a Disentangled Representation learning-Generative Adversarial Network (DR-GAN) with three distinct novelties. First, the encoder-decoder structure of the generator enables DR-GAN to learn a representation that is both generative and discriminative, which can be used for face image synthesis and pose-invariant face recognition. Second, this representation is explicitly disentangled from other face variations such as pose, through the pose code provided to the decoder and pose estimation in the discriminator. Third, DR-GAN can take one or multiple images as the input, and generate one unified identity representation along with an arbitrary number of synthetic face images. Extensive quantitative and qualitative evaluation on a number of controlled and in-the-wild databases demonstrate the superiority of DR-GAN over the state of the art in both learning representations and rotating large-pose face images.

**EPRO BIO-  
INFO - 023**

**Robust Face Image Super-Resolution via Joint Learning of Subdivided Contextual Model**

In this paper, we focus on restoring high-resolution facial images under noisy low-resolution scenarios. This problem is a challenging problem as the most important structures and details of captured facial images are missing. To address this problem, we propose a novel local patch-based Face Super-Resolution (FSR) method via the joint learning of the contextual model. The contextual model is based on the topology consists of contextual sub-patches, which provide more useful structural information than the commonly used local contextual structures due to the finer patch size. In this way, the contextual models are able to recover the missing local structures in target patches. In order to further strengthen the structural compensation function of contextual topology, we introduce the recognition feature as additional regularity. Based on the contextual model, we formulate the super-resolved procedure as a contextual joint representation with respect to the target patch and its adjacent patches. The high-resolution image is obtained by weighting contextual estimations. Both quantitative and qualitative validation shows that the proposed method performs favorably against state-of-the-art algorithms.

**EPRO BIO-  
INFO - 024**

**Tattoo Image Search at Scale: Joint Detection and Compact Representation Learning**

The explosive growth of digital images in video surveillance and social media has led to the significant need for efficient search of persons of interest in law enforcement and forensic applications. Despite tremendous progress in primary biometric traits (e.g., face and fingerprint) based person identification, a single biometric trait alone cannot meet the desired recognition accuracy in forensic scenarios. Tattoos, as one of the important soft biometric traits, have been found to be valuable for assisting in person identification. However, tattoo search in a large collection of unconstrained images remains a difficult problem, and existing tattoo search methods mainly focus on matching cropped tattoos, which is different from real application scenarios. To close the gap, we propose an efficient tattoo search approach that is able to learn tattoo detection and compact representation jointly in a single convolutional neural network (CNN) via multi-task learning. While the features in the backbone network are shared by both tattoo detection and compact representation learning, individual latent layers of each sub-network optimize the shared features toward the detection and feature learning tasks, respectively. We resolve the small batch size issue inside the joint tattoo detection and compact representation learning network via random image stitch and preceding feature buffering. We evaluate the proposed tattoo search system using multiple public-domain tattoo benchmarks, and a gallery set with about 300K distracter tattoo images compiled from these datasets and images from the Internet.

Action recognition has become a rapidly developing research field. But with the increasing demand for large-scale data, the need of hand-annotated data for training becomes more and more impractical. One way to avoid frame-based annotation is the use of action ordering information to learn the respective action classes. We propose a hierarchical approach to address the problem of weakly supervised learning from ordered action labels by structuring recognition in a coarse-to-fine manner. Given a set of videos and an ordered list of occurring actions, we infer the start and end of the related action classes and train the respective action classifiers without any need for hand labeled frame boundaries by combining a framewise RNN model with a coarse probabilistic inference. This combination allows for the temporal alignment of long sequences and thus, for an iterative training of both elements. While this system alone already generates good results, we show that the performance can be improved by approximating the number of subactions to the characteristics of different action classes as well as by the introduction of a regularizing length prior. The proposed system is evaluated on two benchmarks, the Breakfast and the Hollywood extended dataset, showing a competitive performance on temporal action segmentation and alignment.



*THANK YOU!*

Elysium PRO



Titles with Abstracts 2019-20