Mitosis Detection in Breast Cancer Histological Images

Organizers

IPAL Laboratory, UMI CNRS, Singapore
Ludovic Roux, Daniel Racoceanu, Nicolas Loménie, Maria Kulikova, Humayun Irshad
TRIBVN Company, Paris, France
Jacques Klossa
Pitié-Salpêtrière Hospital, Paris, France
Prof. Frédérique Capron, Dr. Catherine Genestie, Gilles Le Naour
CIALAB, The Ohio State University, USA
Metin Gurcan

Website

http://ipal.cnrs.fr/ICPR2012/

Abstract

Nottingham Grading System is an international grading system for breast cancer recommended by the World Health Organization. It is derived from the assessment of three morphological features: architecture (tubule formation), nuclear atypia (also known as nuclear polymorphism) and mitotic count. Several studies on automatic tools to process digitized slides have been reported focusing mainly on nuclei or tubule detection. Mitosis detection is a challenging problem and has not been addressed well in the literature. This contest is a unique initiative that provides a database of digitized microscopic images where mitosis have been annotated. As of July 15th, 2012, we had 129 companies, research institutes or universities in 40 countries that requested to register to the contest. This proves there is a real interest and a need for such a database.

Mitotic count is an important parameter, possibly the most important one, in cancer grading in general, and in breast cancer grading in particular, as it gives an evaluation of the aggressiveness of the tumour. Detection of mitosis is a very challenging task since mitosis are small objects with a large variety of shape configurations. The four main phases of a mitosis are prophase, metaphase, anaphase and telophase. The shape and texture of the nucleus are very different from a phase of the mitosis to another phase, and also within the same phase, which makes the challenge very difficult.

The competition consists in being able to tell what is the mitotic count on an image. Different types of images are provided: Two different slide scanners having different resolution produce RGB images, and a multi-spectral microscope produce images in 10 different spectral bands and 17 layers Z-stack. The contestants can take advantage of using the information of some of the spectral bands which may be more discriminating for the detection of mitosis, or to concentrate only on RGB images.

People tracking in wide baseline camera network

Organizers

Gustavo Fernandez, Austria Institute of Technology, AT (gustavo.fernandez@ait.ac.at)
Helmut Grabner, ETH-Zurich, CH (grabner@vision.ee.ethz.ch)
Fatih Porikli, Mitsubishi Electric Research Laboratories, USA (fatih@merl.com)
Sergio Velastin, Kingston University London, UK (Sergio.Velastin@kingston.ac.uk)

Website

http://www.wide-baseline-camera-network-contest.org/

Abstract

a. General description of the problem

With the increase of security level at public infrastructures, surveillance camera networks having wide baseline setting (including the case of non overlapping fields of view) have received more attention. Driven by cost, surveillance cameras are in most cases distributed and placed in such a way that an observed area is maximised resulting in a very tiny or no overlap of camera fields of view. In some cases, the fields of view overlap but the camera viewpoints are such that the people are seen from totally different views. Many efforts have been done on the Computer Vision area to develop algorithms to track people, mainly using a single camera. However, aforementioned camera systems using Computer Vision algorithms still remain a challenge. We are interested in going a step further and evaluate the performance of possible solutions for tracking people in such camera systems.

b. Datasets to be used

Three different video datasets, capturing various real-world scenarios, are provided. Each dataset was generated using either 4 or 6 cameras.

1. PETS 2006 (http://www.cvg.rdg.ac.uk/PETS2006/data.html)

   4 cameras, all fields of view are overlapping, wide baselines, occlusions of people by themselves and large distance between the camera and people.

2. CAT (http://www.cat-project.at/)

   4 cameras, 4 persons walking around simultaneously, all cameras have at most slightly overlapping fields of view, non-overlapping fields of view between some cameras and different illumination conditions.

3. Techgate

   6 cameras, mainly non-overlapping fields of view, slightly overlapping fields of view between some cameras and occlusions of people by themselves.

c. Competition task

The competition aims accurate and consistent tracking of a single person and multiple people through a given multi-camera system.

Due to little or no overlap between camera field-of-views, a major challenge is the exchange and update of visual information from one camera to next. Another challenge is the selection of suitable image features for these tasks. The spatial accuracy and label consistence of the tracking results will be the main determinants of the contest.

Objects can be initialized manually or automatically. After the initialization stage, the tracking process is required to be fully automatic. This means, no additional user interaction or user provided information, e.g. new manual selections, can be provided during the tracking.

Human activities recognition and localization competition

Organizers

Christian Wolf	INSA-Lyon, France
Julien Mille	Université Claude Bernard Lyon 1, France
Eric Lombardi	CNRS, France
Oya Celiktutan	Bogazici University, Istanbul, Turkey
Mingyuan Jiu	INSA-Lyon, France
Moez Baccouche	INSA-Lyon, France
Emmanuel Dellandréa	Ecole Centrale de Lyon, France
Charles-Edmond Bichot	Ecole Centrale de Lyon, France
Christophe Garcia	INSA-Lyon, France
Bülent Sankur	Bogazici University, Istanbul, Turkey

Website

http://liris.cnrs.fr/harl2012

Abstract

The tasks of the HARL competition focus on the localisation and recognition of complex human behavior from video sequences, possibly involving several people present at the same time. An emphasis is set on actions involving several interacting people and on human-object interactions. The full data set contains 828 actions of 10 different classes performed by 21 different people. In order to make the dataset more challenging than previous datasets, the actions are less focused on low level characteristics and more defined by semantics and context: 1. Discussion; 2. Giving an item; 3. Picking up or droping item; 4. Entering/leaving room; 5. Trying to enter/leave room; 6. Unlocking and entering; 7. Leaving baggage unattended; 8. Handshaking; 9. Typing on keyboard; 10. Talking on a telephone.

The LIRIS dataset specifically acquired for this competition has been shot with two different cameras: a moving camera mounted on a mobile robot delivering grayscale videos in VGA resolution and depth images from a consumer depth camera (Kinect); and a consumer camcorder delivering color videos in DVD resolution.

The data have been annotated by marking labeled bounding boxes for each frame of each action. The evaluation metric has been designed to measure the degree of similarity between the two ground truth and the detection result, with two goals in mind: i) the metric provides a quantitative evaluation, it intuitively tells how many actions have been detected correctly, and how many false alarms have been created; ii) the metric also provides a qualitative evaluation, it gives an easy interpretation of the detection quality.

The competing methods used varying techniques, ranging from segmentation free methods based on space-time interest points and bags of words models, to methods involving people detectors, skin classification and human pose estimation.

More information on the dataset and on the evaluation method can be found in the following reference

[1] C. Wolf, J. Mille, E. Lombardi, O. Celiktutan, M. Jiu, M. Baccouche, E. Dellandr´ea, C.-E. Bichot, C. Garcia, and B. Sankur. The liris human activities dataset and the icpr 2012 human activities recognition and localization competition. Technical Report LIRIS RR-2012-004, Laboratoire d'Informatique en Images et Syst`emes d'Information, INSA de Lyon, France, 2012.

CHALEARN Gesture Challenge

Organizers

Isabelle Guyon, ChaLearn, USA
Vassilis Athitsos, University of Texas at Arlington, USA
Alex Kipman, Microsoft, USA

Website

http://gesture.chalearn.org

Abstract

This is a challenge on gesture and sign language recognition using a Kinect camera. Kinect is revolutionizing the field of gesture recognition by providing an affordable 3D camera, which records both an RGB image and a depth image (using an infrared sensor). The challenge focuses on hand gestures. Applications include man-machine communication, translating sign languages for the deaf, video surveillance, and computer gaming. Every application needs a specialized gesture vocabulary. If we want gesture recognition to become part of everyday life, we need gesture recognition machines, which easily get tailored to new gesture vocabularies. This is why the focus of the challenge is "one-shot-learning" of gestures, which means learning to recognize new categories of gestures from a single video clip of each gesture. The gestures are drawn from a small vocabulary of gestures, generally related to a particular task, for instance, hand signals used by divers, finger codes to represent numerals, signals used by referees, or marchalling signals to guide vehicles or aircrafts.

The competition had two identical rounds. This is the final round. Like in round 1, there was a development phase (May 7, 2012 to September 6, 2012) and a final evaluation phase (September 7, 2012 to September 10. 2012). During the development phase the competitors created a learning system capable of learning from a single training example a gesture classification problem. They practiced with development data (a large database of 50,000 labeled gestures) and submitted predictions on-line on validation data to get immediate feed-back on a leaderboard. Towards the end of the development phase, the participants had the opportunity to submit their code for verification purpose. During the final evaluation phase, they made predictions on the new final evaluation data revealed at the end of the development phase. The participants had then 4 days to train their systems and upload their predictions. The results of the top ranking participants who submitted their code were verified by the organizers.

HEp-2 Cells Classification

Chair

Gennaro Percannella (pergen@unisa.it, University of Salerno, Italy)

Expert committee members

Pasquale Foggia (pfoggia@unisa.it, University of Salerno, Italy)
Paolo Soda (p.soda@unicampus.it, University "Campus Bio-Medico" of Roma, Italy)

Website

http://mivia.unisa.it/hep2contest

Abstract

The contest focuses on Indirect Immunofluorescence (IIF), a method suited to reveal the presence of autoimmune diseases by searching antibodies in the patient serum for diagnostic purposes. Due to its effectiveness, in the recent years we have assisted to a growing demand of diagnostic tests for systemic autoimmune diseases. Unfortunately, today IIF is still a subjective method too dependent on the experience and expertise of the physician. As a consequence, there is a strong demand for a complete automation of the procedure. In fact, a fully automated IIF procedure would guarantee easier and faster result reporting, increased test repeatability and lower costs.

In the last years some research groups have proposed different algorithms for the analysis of IIF images. However, the validation of the proposed methods has been carried out on small and private datasets. This competition aims to bring researchers interested in performance evaluation of algorithms for IIF images analysis on the same dataset. Through this competition it will be possible to compare different solutions on a large and significant set of real data. Furthermore, this initiative is also intended to increase the interest of the scientific community toward this applicative area.

The competition concerns the recognition of the cell patterns in HEp-2 images. In this regard the contest organizers provided to the participants the original images together with the cells already segmented by specialists.

The dataset is constituted by 28 images, which, as a whole, contain 1457 cells, almost equally distributed among the different patterns. The participants received the cells belonging to 14 out of the 28 images, to be used as the training set. The subdivision in training and test sets was done maintaining approximately the same pattern distribution over the two datasets. The test set was not disclosed to the participants.

The participants were required to design and implement a pattern recognition system able to classify the pre-segmented cells belonging to HEp-2 images in one of the following pattern classes: homogeneous, fine speckled, coarse speckled, nucleolar, cytoplasmatic and centromere. The application obtaining the highest value of the accuracy in cell classification over the private test set is declared as the winner.

The initiative got a large interest from the scientific community: 135 registrations to the contest website and 28 executables submitted for testing by research groups from almost all the continents.

Kitchen Scene Context based Gesture Recognition

Organizers

Atsushi Shimada (Kyushu University, Japan)
Kazuaki Kondo (Kyoto University, Japan)
Daisuke Deguchi (Nagoya University, Japan)
Géraldine Morin (IRIT-ENSEEIHT, France)
Helman Stern (Ben Gurion University of the Negev, Israel)

Website

http://www.murase.m.is.nagoya-u.ac.jp/KSCGR/

Abstract

A gesture has strong relationship not only with an actor's motion, but also with the situation surrounding the actor. Traditionally, most gesture recognition methods focus on motion features only, and assign the gesture label based on the discriminative analysis. There are, however, many gestures which cannot always be uniquely determined by using motion features alone. For instance, considering gestures in a kitchen, "mixing something" and "sauteing something" would be done by similar motion sequences of a hand, which moves around in a circular motion. If the cook uses a bowl, the gesture label (we call it "cooking motion" in the contest) should be "mixing something in a bowl". Likewise, if he/she takes a frying pan, the label would be regarded as "sauteing something in a frying pan". Thus, the gesture label should be estimated by consideration of both the motion features and the context situation in which is it performed.

The contestants are expected to evaluate the human gestures in a kitchen from continuous video sequences. Five-types of cooking menus using the ingredients of egg, milk, and ham are considered in the contest. The candidates of the cooking menus are "ham and eggs", "omelet", "scrambled egg", "boiled egg" and "Kinshi-tamago". An actor cooks one of cooking menus with above ingredients. Each video contains several cooking motions: pouring, mixing, boiling, sauteing and so on. Using the testing videos, contestants have to establish the relationships between motion features and scene features (i.e. scene contexts) and achieve scene context based cooking motion recognition by assigning a correct cooking motion label to each video frame.

There are five candidate cooking menus. Each menu is cooked by five different actors; that is, five cooking scenes are available for each menu. The scene is captured by a Kinect sensor providing synchronized color and depth image sequences. Each of the videos is from 5 to10 minute long containing 2,000 to 12,000 frames. A cooking motion label is assigned to each frame by the organizers, indicating the type of gesture performed by the actors. These labels are regarded as ground truth in the contest.

Contestants must evaluate the testing video sequence in order to estimate the cooking motion in each frame. Of course, scene contexts including ingredients, cooking tools and motions should be extracted from the training video sequences in advance. Using the scene context is not an indispensable condition, but the organizers strongly recommend utilizing it.