

Identification and Retrieval of prostate cancer cases using a content based search tool.

Background: Large amounts of histopathology digital images have become available over the past years. This allows to algorithmically analyse them and makes them available for visual image retrieval. However, retrieving histopathology images from varied sources, such as online scientific literature and teaching files, is difficult due to their heterogeneity in magnification and a lack of metadata. We present a retrieval system and a scale detection method that allow browsing in heterogeneous histopathology datasets.

Methods: A whole slide image viewer integrates the visual image retrieval system for digital pathology. The dataset includes whole slide images and histopathology images automatically extracted from the biomedical open access literature of Pub Med Central. The scale detection algorithms are based on artificial neural networks.

Results: The retrieval system allows searching for regions that are visually similar to manually delineated regions of interest and identify similar cases. Scale detection allows to identify the scale of the images and define adapted features for retrieval and classification, even if the magnification level is unknown. The first version of the retrieval system has been tested by pathologists, who highlighted its capabilities and suggested possible ways to improve it by adding a larger number of cases.

Conclusions: A visual histopathology retrieval system can enhance the practice of pathologists allowing them to easily retrieve similar regions in image or similar cases in proprietary datasets. The scale detection improves the retrieval performance on heterogeneous datasets and allows pathologists to retrieve images from scientific publications and possibly teaching files and books.