IMPROVED STATISTICAL RECOGNITION ALGORITHMS FOR OIL PALM RIPENESS IDENTIFICATION

FATMA SUSILAWATI BINTI MOHAMAD

UNIVERSITI TEKNOLOGI MALAYSIA

IMPROVED STATISTICAL RECOGNITION ALGORITHMS FOR OIL PALM
RIPENESS IDENTIFICATION

FATMA SUSILAWATI BINTI MOHAMAD

A thesis submitted in the fulfilment of the

requirements for the award of the degree of

Doctor of Philosophy (Computer Science)

Faculty of Computer and Information System

Universiti Teknologi Malaysia

NOVEMBER 2012

To my beloved mother, husband, kids and family members

ACKNOWLEDGEMENT

ABSTRACT

Awareness for high quality crude oil is crucial in oil palm production. Proper grading process is important to ensure only the ripe fruits are taken into consideration for the maximum level of oil content. Currently, researchers focus mainly on providing an automatic grading system using various techniques such as producing digital numbers, oil palm colorimeter, photogrammetric grading, fuzzy or neuro-fuzzy technique and so on. Even though some of them have more than 85% accuracy, it is only valid in controlled environment. However, when they are applied in real situation with uncontrolled environment, the accuracy can drop to less than 50%. So far, there is limited study on suitable colour model conducted on oil palm ripeness identification. Most researchers use RGB colour model to determine an oil palm ripeness. This research looks into the suitability and performance of HSV colour model in classifying an oil palm ripeness. Distance Measurement and Linear Discriminant Analysis are chosen as methods to classify an oil palm ripeness in this study. Histogram is used as a feature vector for feature extraction method while colour as a feature to be analysed. Images of oil palm were captured by an expert in the form of JPEG images. Preprocessing is then performed to remove noise and background from the images. Subsequently, images are transformed into histogram and mean value are extracted. Selected Distance Measurement such as Euclidean Distance, Nearest Neighbour, Furthest Neighbour and Mean Distance are then used for feature matching process. An Oil Palm Ripeness Identification algorithm is proposed, wherein an elimination technique is also introduced in the process. In addition, a Multiple Features Technique is also proposed to find the best feature which brings a very good recognition rate for selected Distance Measurement. The results show that 98% accuracy have been obtained in comparison with other researchers' work.

ABSTRAK

Kesedaran untuk menghasilkan minyak mentah berkualiti tinggi adalah penting dalam pengeluaran minyak sawit. Proses penggredan yang betul adalah penting untuk memastikan hanya buah masak diambil kira untuk tahap maksimum kandungan minyak. Pada masa ini, penyelidik memberi tumpuan terutamanya dalam menyediakan sistem penggredan automatik menggunakan pelbagai teknik seperti menghasilkan nombor digital kolorimeter, minyak sawit, penggredan fotogrametri, teknik kabur atau kabur neuro dan sebagainya. Walaupun sebahagian daripada mereka mempunyai lebih daripada 85% ketepatan, ia hanya sah dalam persekitaran terkawal. Walau bagaimanapun, apabila ia digunakan dalam situasi sebenar dengan persekitaran yang tidak terkawal, ketepatan boleh jatuh kepada kurang daripada 50%. Setakat ini, kajian yang dijalankan terhadap model warna yang sesuai adalah terhad ke atas pengenalan kematangan kelapa sawit. Kebanyakan penyelidik menggunakan model warna RGB untuk menentukan kematangan kelapa sawit. Kajian ini melihat kepada kesesuaian dan prestasi model warna HSV dalam mengklasifikasikan kematangan kelapa sawit. Jarak Pengukuran dan Analisis pembezalayan Linear dipilih sebagai kaedah untuk mengklasifikasikan kematangan kelapa sawit dalam kajian ini. Histogram digunakan sebagai ciri vektor bagi kaedah pengekstrakan ciri manakala warna adalah ciri yang akan dianalisis. Gambar kelapa sawit telah ditangkap oleh seorang pakar dalam bentuk imej JPEG. Prapemprosesan kemudian dilakukan untuk menghilangkan bunyi dan latar belakang dari imej. Selepas itu, imej diubah ke dalam bentuk histogram dan nilai purata diekstrak. Jarak Pengukuran terpilih seperti Jarak Euclid, Jarak Jiran Terhampir, Jarak Jiran Paling Jauh dan Jarak Purata kemudiannya digunakan untuk pencarian ciri sepadan. Satu algoritma Pengenalan Kematangan Kelapa Sawit dicadangkan dimana teknik penghapusan juga diperkenalkan dalam proses. Di samping itu, Teknik Pelbagai Ciri juga dicadangkan untuk mencari ciri terbaik yang membawa pengiktirafan kadar pengecaman yang amat baik bagi Pengukuran Jarak terpilih. Keputusan menunjukkan bahawa ketepatan 98% telah diperolehi berbanding dengan kerja penyelidik yang lain.