Ministry of Higher Education and
Scientific Research
University of Mosul
College of Computer Science and
Mathematics
Department of Computer Science

Specified Multi Object Detection And Tracking In Video File Based On Deep Learning

**A Thesis Submitted to the Council of the College of
Computer Science and Mathematics
University of Mosul
as a Partial Fulfillment of Requirements
for the Degree of Doctor of Philosophy
in
Computer Science**

By

Younis Abbas Younis Al Arbo

Supervised by

Prof. Dr. Khalil Ibrahim Al Saif

2022 A.D. **1443 A.H.**

Abstract

This research designs and proposes a new “Multi-object Tracking” (MOT) method for pedestrian tracking and recognition. To work towards this aim, the study modifies the already existing method of detection by tracking through the use of descriptors extracted from the convolutional filters of “You Only Look Once (YOLOv3)” neural network called m-yolo. As a result, the features detected and processed during the detection phase are exploited to produce efficient and robust descriptors. Consequently, the median function is applied on the last three descriptors to resist noise and/or ignore anomalous values which could be predicted by the algorithm. For matching, the “Euclidean distance” was used forthe purpose of calculating the matching between descriptors to get good tracking performance. The Compared to current approaches, this m-yolo is more effective, since it produces better predictions and uses fewer computations. The results record 98.63% for average precision, 79.6% for identity F1-score, and 62.9% for “Multi-Object Tracking Accuracy” (MOTA).

The same descriptors used with YOLOv3 are further embedded with “Faster Region Convolution Neural Network”. This embedding allows the proposed method to directly output a descriptor for each object detected by Faster R-CNN. Each of these descriptors connect a deep feature extracted from “Region Proposal Network” (RPN) and a visual feature extracted from Pre- trained CNN. The use of these features allows the proposed method to acquire accurate values rapidly, since it is used three convolution layers to gain only distinctive feature that represent the target

Consequently, the descriptors are forwarded to deep sort, The detection algorithm with m-Faster R-CNN is evaluated using the “Penn-Fudan database” and has recorded 98.45% for average precision, while the tracking algorithm is evaluated using the “MOT16 challenge benchmark pedestrian dataset” and has recorded 80.4% for identity F1-score, and 62.5% for MOTA.

الملخص

يقترح هذا البحث طريقة جديدة "لتتبع الأجسام المتعددة" (MOT) لتتبع المشاة والتعرف عليها. للعمل على تحقيق هذا الهدف ، تعدل الدراسة طريقة التتبع عن طريق الكشف الموجودة بالفعل ، بالاعتماد على الواصفات المستخرجة باستخدام المرشحات التلافيفية للشبكة العصبيةYOLOv3 . نظرًا لاكتشاف هذه الميزات ومعالجتها أثناء مرحلة الاكتشاف ، فقد استغلت الطريقة المقترحة هذه الميزات لإنتاج واصفات فعالة وقوية ثم تم تطبيق الداله الوسيطة لاخر ثلاثة واصفات لغرض مقاومة الضوضاء او القيمه المتطرفة التي من الممكن ان يتنبئ بها الخوارزمية اما بالنسبة للمطابقة تم استخدام المسافة الاقليدية لغرض حساب التطابق بين الواصفات للحصول على اداء تتبع جيد. أظهرت الطريقة المقترحة أداءً أفضل ، مقارنةً بأحدث الأساليب ، من خلال إنتاج تنبؤات أفضل باستخدام حسابات أقل. حيث تم الحصول على متوسط الدقه بمقدار98.63% وتمكنت الخوارزمية من تحقيق IDF1 بمقدار 79.6٪ بينا كانت نتيجة MOTA تقدر ب 62.9٪.

تم تضمين نفس الواصفات المستخدمة عند YOLOv3 مع Faster R-CNN يسمح هذا التضمين للطريقة المقترحة بإخراج واصف مباشرة لكل كائن تم اكتشافه بواسطة Faster R-CNN ، يربط كل من هذه الواصفات المميزات العميقة المستخرجة من "شبكة اقتراح المنطقة" (RPN) والمميزات البصرية المستخرجة من شبكة CNN المدربة مسبقًا.. يسمح استخدام هذه الميزات للطريقة المقترحة بإخراج قيم دقيقة بسرعة ، حيث تم استخدام ثلاث طبقات التفاف لاكتساب السمة المميزة فقط التي تمثل الهدف. يتم إعادة توجيه الواصفات التي تم جمعها من الطريقة المقترحة إلى الفرز العميق. ثم ، بالنسبة للإطارات الإضافية ، يعتمد الفرز العميق على مرشح kalman للتنبؤ بالهدف في الإطار التالي من خلال الاعتماد على الإطارات السابقة وسرعة الهدف. تم تقييم خوارزمية الاكتشاف باستخدام قاعدة بيانات Penn-Fudan وتمكنت من تحقيق 98.45% لمتوسط الدقة بينما يتم تقييم خوارزمية التتبع باستخدام مجموعة بيانات المشاة المعيارية لتحدي MOT16 وتمكنت من تحقيق IDF1 بنسبة 80.4٪ و MOTA بنسبة 62.5٪.