



الجمهورية العربية السورية

جامعة دمشق

كلية الهندسة الميكانيكية والكهربائية

قسم هندسة الحواسيب والأتمتة

تصميم نظام لكشف السقوط المؤذي للإنسان

بكاميرات المراقبة

أطروحة مقدمة لنيل درجة الدكتوراه في هندسة التحكم والأتمتة

إعداد

م. عمار محمد ياسر الدقاق

إشراف

المشرف المشارك

د. م. رشا كاظم مسعود

أستاذ في قسم الهندسة الطبية

المشرف الرئيسي

د. م. محمد مازن المحاييري

أستاذ مساعد في قسم هندسة الحواسيب والأتمتة

2023 م

لجنة الحكم

قرار مجلس البحث العلمي والدراسات العليا رقم /188/ المتخذ بالجلسة رقم /2/ تاريخ 2023/10/2 بالموافقة على تأليف لجنة الحكم على رسالة الدكتوراه في قسم هندسة الحواسيب والأتمتة التي أعدها الطالب **عمار الدقاق** بعنوان: ((تصميم نظام لكشف السقوط المؤذي للإنسان بكاميرات المراقبة)) بكلية الهندسة الميكانيكية والكهربائية

من السادة الأساتذة:

كلية الهندسة الميكانيكية والكهربائية

الأستاذ في قسم هندسة الحواسيب والأتمتة

د. سمير كرمان

الاختصاص: عناصر وبنية الحاسبات الالكترونية

جامعة دمشق

كلية الهندسة الميكانيكية والكهربائية

الأستاذ المساعد في قسم هندسة الحواسيب والأتمتة

د. محمد مازن المحاييري

الاختصاص: هندسة تنظيم الحاسبات وبنيتها

جامعة دمشق

كلية الهندسة الميكانيكية والكهربائية

الأستاذ المساعد في قسم الهندسة الطبية

د. مصطفى الموالي

الاختصاص: ميكانيك حيوي وأطراف صناعية وأجهزة تقييمية

جامعة دمشق

كلية الهندسة الميكانيكية والكهربائية

الأستاذ المساعد في قسم هندسة الحواسيب والأتمتة

د. مأمون يونس

الاختصاص: هندسة الحواسيب والمجمعات والمنظومات والشبكات الحاسوبية

جامعة دمشق

كلية الهندسة الميكانيكية والكهربائية

المدرس في قسم هندسة الحواسيب والأتمتة

د. هيام خدام

الاختصاص: التحكم الرقمي

جامعة دمشق

الإهداء

إلى والدي الكريمين اللذين ربياني وحثاني على محبة العلم...
إلى زوجتي العزيزة التي تؤازرني وتشاركني مصاعب الحياة...
إلى عصفوري الصغيرين - ياسر وجمان - اللذين أعتبرهما عيناى اللتين أرى بهما...
إلى أخي العزيز وأقاربي...
إلى أصدقائي وزملائي...
إلى العالم العربي متمنياً أن يكون هذا العمل لبنة ولو صغيرة في بناء صرح العلم فيه...

الشكر والتقدير

أود أن أتقدّم بجزيل الشكر والتقدير لكل من

الأستاذ الدكتور المهندس محمد مازن المحاييري والأستاذة الدكتورة رشا كاظم مسعود

الذين لم يتوانيا عن تشجيعي وتوجيهي وتزويدي بالمراجع والنصائح القيمة...
كما أود أن أتوجه بالشكر والامتنان لأعضاء لجنة الحكم:

د. سمير كرمان

د. مازن المحاييري

د. مصطفى الموالي

د. مأمون يونس

د. هيام خدام

على ملاحظاتهم القيّمة ونقدتهم البناء الذي ساهم في رفع سويّة هذا العمل.

فهرس المحتويات

الفصل	الصفحة
لجنة الحكم.....	2
الإهداء.....	3
الشكر والتقدير.....	4
فهرس المحتويات.....	5
فهرس الأشكال.....	9
فهرس الجداول.....	12
مسرد المصطلحات والاختصارات.....	13
الملخص.....	20
المقدمة.....	22
أ. مسوغات البحث وأهدافه.....	23
ب. مناهج البحث وحدوده وأدواته.....	24
ت. موجز محتويات الفصول القادمة.....	24
1- الفصل الأول شبكات التعلم العميق.....	26
1-1- تمهيد.....	26
1-2- الشبكات العصبونية الاصطناعية.....	27
1-3-1- تعليم الشبكة العصبونية.....	30
1-3-1- التعلم الخاضع للإشراف Supervised Learning.....	30
1-3-2- التعلم غير الخاضع للإشراف Unsupervised Learning.....	31
1-3-3- التعلم المعزز Reinforcement Learning.....	32

32	4-1- شبكات التعلّم العميق
32	1-4-1 الشبكات العصبونية الالتفافية
36	2-4-1 طبقة الالتفاف أو الطي Convolution
38	3-4-1 طبقة التجميع pooling
43	4-4-1 نقل التعلم Transfer Learning
43	5-1 شبكات الذاكرة القصيرة طويلة الأمد (LONG SHORT-TERM MEMORY LSTM)
47	6-1 الشبكات الهجينة CNN-LSTM
49	7-1 معايير تقييم عملية تعليم الشبكة
51	8-1 الخلاصة
52	2- الفصل الثاني صورة تاريخ الحركة MOTION HISTORY IMAGE
52	1-2 تمهيد
53	2-2 صورة تاريخ الحركة MHI وصورة طاقة الحركة MEI
56	1-2-2 تأثير المعاملان τ و δ
58	2-2-2 تأثير تابع التحديث Ψ ومعامل تحسس الحركة ξ
61	3-2 قصور طريقة MHI
64	4-2 منهجيات استخدمت فيها MHI
65	5-2 مناهج أخرى مطورة لطريقة MHI
71	6-2 الخلاصة
72	3- الفصل الثالث الدراسة المرجعية
72	1-3 تمهيد
72	2-3 أنواع طرق كشف السقوط
72	1-2-3 طرق تعتمد الحساسات المحيطية
73	2-2-3 طرق تعتمد الأجهزة الملاصقة للجسم

75	3-2-3 طرق تعتمد الإبصار الحاسوبي.....
86	3-3 الخلاصة.....
88	4- الفصل الرابع تطوير شبكة CNN-LSTM لكشف السقوط.....
88	4-1 تمهيد.....
89	4-2 بنية الشبكة المقترحة.....
92	4-3 مرحلة التدريب والاختبار.....
92	4-3-1 معطيات التدريب.....
97	4-3-2 مرحلة التدريب.....
99	4-4 النتائج والمناقشة.....
104	4-5 الخلاصة.....
106	5- الفصل الخامس استخدام صورة الحركة STTFI المقترحة في منظومة كشف السقوط.....
106	5-1 تمهيد.....
107	5-2 صورة أمامية مكدسة بعلامات زمنية (STTFI) STACKED TIME-TAGGED FOREGROUND IMAGE
114	5-3 منظومة كشف السقوط المقترحة باستخدام صور STTFI.....
114	5-3-1 استحصال الأطر من الفيديو.....
115	5-3-2 كشف هيئة الإنسان Person detection.....
116	5-3-3 استخلاص صورة STTFI.....
116	5-3-4 الشبكة العصبونية الالتفافية CNN.....
117	5-4 مراحل تحضير المعطيات والتدريب.....
117	5-4-1 معطيات التدريب.....
119	5-4-2 تحضير المعطيات.....
120	5-4-3 مرحلة التدريب.....
120	5-5 النتائج والمناقشة.....

126	6-5- الخلاصة
128	6- الخاتمة
129	6-1- الاستنتاجات
130	6-2- الأفاق المستقبلية
132	المراجع
145	ABSTRACT

فهرس الأشكال

الصفحة	الشكل
27.....	الشكل 1—1 (A) شبكة عصبونية أمامية وحيدة الطبقة، (B) شبكة عصبونية أمامية متعددة الطبقات [41].....
28.....	الشكل 1—2 مخطط يبين كيفية حساب خرج عقدة أو عصبون من الشبكة العصبونية الاصطناعية [41].....
29.....	الشكل 1—3 شبكة تكرارية بطبقة عصبونات خفية [40].....
33.....	الشكل 1—4 شبكة LENET العصبونية الالتفافية لتمييز الأحرف [48].....
34.....	الشكل 1—5 شبكة ALEXNET العصبونية الالتفافية [39].....
36.....	الشكل 1—6 تابع التنشيط RELU المستخدم في الشبكات العصبونية الالتفافية [50].....
38.....	الشكل 1—7 رسم توضيحي يبين كيفية حساب خرج طبقة الطي في الشبكات العصبونية الالتفافية [49].....
40.....	الشكل 1—8 رسم توضيحي يبين كيف يحسب خرج طبقة MAXPOOLING [51].....
41.....	الشكل 1—9 يوضح ناتج تطبيق المرشح التجميعي ذو المتوسط العام على خرائط السمات [52].....
42.....	الشكل 1—10 بني عدة شبكات التفافية مختلفة [53].....
44.....	الشكل 1—11 بنية الشبكة RNN الأساسية وحالتها أثناء مرور الزمن [57].....
45.....	الشكل 1—12 خلية LSTM العصبونية التكرارية [57].....
47.....	الشكل 1—13 البنى المختلفة لشبكات RNN [57].....
48.....	الشكل 1—14 شبكة ثنائية السيل لتصنيف الفيديو [58].....
49.....	الشكل 1—15 بنية الشبكة الهجينة CNN-LSTM [60].....
51.....	الشكل 1—16 رسم يبين مصفوفة الالتباس [67].....
54.....	الشكل 1—2 A- صور شخص يلوح بذراعيه، B- صورة MHI [4].....
	الشكل 2—2 الصف الأول الأعلى من اليسار يمثل أربعة إطارات من فيديو لرجل ينحني رافعاً يده اليسرى، أما الصورة الخامسة فهي صورة MHI الممثلة لتلك الحركة، والسادسة هي صورة MEI، أما الصف الثاني يمثل أربعة إطارات لرجل ينحني نزولاً باستخدام يده اليمنى، والصورة الخامسة هي صورة MHI لهذه الحركة، أما السادسة فهي صورة MEI [69].....
57.....	الشكل 2—3 صور MHI لذات المشهد ولكن بقيم T مختلفة [69].....

- الشكل 2—4 صور MHI لنشاط الانحناء ولكن بقيم مختلفة ل δ [69]..... 58
- الشكل 2—5 أهمية اختيار معامل التحسس أو العتبة بالنسبة لتابع التحديث [69]..... 59
- الشكل 2—6 صور MHI لفتاة الأيروبيك تحرك جسمها ببطء [4]..... 59
- الشكل 2—7 صور كل من MHI و OF-MHI لثلاثة نشاطات مختلفة [80]..... 61
- الشكل 2—8 العامودان على اليسار لشخص يبتعد عن الكاميرا مع صورتي MEI بينما على اليمين لشخص يقترب من الكاميرا مع صورتي MEI، حيث تفشل MHI و MEI في تمييز النشاطين [69]..... 63
- الشكل 2—9 شخص يقوم بالقرصاء ثم بعد ذلك يعود إلى الوقوف وصور MHI المرافقة [69]..... 63
- الشكل 2—10 يظهر رجلاً يرفع يده مع بقاء جسده ساكناً، كما يظهر صورة TMHI المقابلة [96]..... 66
- الشكل 2—11 عدة إطارات من مشية شخص مع صورة MSI المقابلة [99]..... 67
- الشكل 2—12 A- صورة FSHI لرجل يمشي، B- صورة BSHI لذات الرجل [100]..... 69
- الشكل 2—13 صور ITMI لمجموعة من النشاطات التي يقوم بها شخص [101]..... 70
- الشكل 2—14 صورة لفتاة تجري مع صورة BMI المقابلة [102]..... 71
- الشكل 4—1 المخطط التدفقي للخوارزمية والشبكة الهجينة المقترحة..... 94
- الشكل 4—2 صور متنوعة من مجموعات معطيات التدريب..... 97
- الشكل 4—3 منحنى الدقة لشبكة ALEXNET المعدلة مقابل عدد دورات التدريب..... 101
- الشكل 4—4 منحنى الدقة لخلية LSTM مقابل عدد دورات التدريب..... 101
- الشكل 5—1 A- مجموعة إطارات لحركة السقوط في حال فقدان الوعي، B- صورة STTFI المستنتجة منها، C- صورة MHI(T=40)..... 108
- الشكل 5—2 صور STTFI و MHI لنشاط الانحناء ثم الاعتدال..... 111
- الشكل 5—3 صور MHI وصورة STTFI لشخص يمشي ثم يجلس..... 112
- الشكل 5—4 صور MHI و STTFI لشخص يسقط إلى الوراء..... 113
- الشكل 5—5 صورة STTFI وصور MHI لشخص يسقط إلى الأمام..... 114
- الشكل 5—6 منظومة كشف السقوط المقترحة باستخدام صورة STTFI..... 115
- الشكل 5—7 صور من مجموعة التدريب FD22 لأحد المتطوعين..... 118

الشكل 5—8 صور من مجموعة التدريب FD22 لأحد المتطوعين 119

فهرس الجداول

الصفحة	الجدول
102	الجدول 4—1 مصفوفة الالتباس CONFUSION MATRIX
102	الجدول 4—2 الحساسية والنوعية لكل من الوضعيات المرحلية
103	الجدول 4—3 مصفوفة الالتباس CONFUSION MATRIX لشبكة LSTM
103	الجدول 4—4 الحساسية والنوعية لشبكة LSTM
121	الجدول 5—1 مصفوفة الالتباس المتعلقة بشبكة ALEXNET المدربة على صور MHI
121	الجدول 5—2 مصفوفة الالتباس المتعلقة بشبكة ALEXNET المدربة على صور STTFI
122	الجدول 5—3 مقارنة بين الطريقة المقترحة وطريقة MHI
123	الجدول 5—4 مصفوفة الالتباس لشبكة ALEXNET المدربة على المجموعات الأربعة باستخدام صور STTFI
123	الجدول 5—5 مقارنة بين الطريقة المقترحة والطرق التي تستخدم الشبكة العصبونية الالتفافية مع صور ملونة أو تدفق بصري
125	الجدول 5—6 مقارنة بين الطريقة المقترحة والطرق التي تستخدم الشبكة العصبونية الالتفافية مع تعدد السيول
126	الجدول 5—7 مقارنة بين الطريقة المقترحة والطرق التي تستخدم شبكة عصبونية عميقة مؤلفة من مرحلتين متتابعيتين

مسرد المصطلحات والاختصارات

Accuracy	الدقة
Activities of Daily Living (ADLs)	النشاطات الحياتية اليومية
Activity Recognition	تمييز النشاط
Average Pooling	التجميع الوسطي
Back Propagation	الانتشار العكسي
Back Propagation Through Time BPTT	الانتشار العكسي خلال الزمن
Batch Size	حجم الدفعة
Binary Motion Image BMI	صورة الحركة الثنائية
Cell	خلية
Classification	تصنيف
Classifier	مصنف
Coefficient	معامل
Condition Random Fields CRF	خوارزمية الحقول العشوائية الشرطية
Confusion Matrix	مصفوفة الالتباس
Convolutional Neural Network (CNN)	شبكة عصبونية التلافيفية
Data Augmentation	زيادة المعطيات
Data Set	مجموعة معطيات تدريب

Depth	عمق
Direct Motion-Based Approaches	طرق تعتمد على الحركة المباشرة
Dropout	الاسقاط العشوائي
Dynamic Image (DI)	صورة ديناميكية
Electroencephalography (EEG)	تخطيط كهربائية الدماغ
Epoch	دورة
Exploding Gradient	التدرج المنفجر
Fallen	سقوط
False Negative FN	سلبي خاطئ
Faster R-CNN	الشبكة التفاضلية الأسرع ذات المناطق
Feature Map	خريطة سمات
Feedforward	أمامية
Forward Propagation	انتشار أمامي
False Positive FP	إيجابي خاطئ
Fully Connected	كاملة الاتصال
Gait Energy Image GEI	صورة طاقة المشي
Gait Recognition	تمييز المشي
Gaussian Mixture Model (GMM)	نموذج المزيج الغوسي
Gaussian Pyramid	الهرم الغوسي
Generalization	التعميم

Generic Human Model Recovery	استعادة النموذج البشري العام
Global Average Pooling	التجميع المتوسط الشامل
Global Positioning System GPS	نظام تحديد المواقع العالمي
Global System for Mobile Communications GSM	النظام العالمي للاتصالات المتنقلة
Gradient Clipping	اقتطاع التدرج
Gradient-Decent Rule	قاعدة الانحدار المتدرج
Graphic Processing Unit GPU	وحدة معالجة رسومية
Ground Truth	القيم الفعلية الصحيحة
Histogram of Oriented Gradients HOG	هيستوغرام الانحدار الموجه
History of Binary Motion Image HBMI	تاريخ صورة الحركة الثنائية
Humoments	شعاع سباعي السمات لا متغير مشتق من العزوم
Hybrid Deep Random Neural Network HDRNN	شبكة عصبونية عشوائية هجينة عميقة
Hybrid Network	شبكة هجينة
Image Captioning	شرح الصورة
ImageNet Large Scale Visual Recognition Challenge (ILSVRC)	تحدي التعرف البصري على نطاق واسع بصور ImageNet
Integrated Time Motion Image (ITMI)	صورة الحركة الزمنية المتكاملة
K-Nearest Neighbor KNN	طريق الـ K جار الأقرب

Kohonen Self-Organizing feature Map SOM	خوارزمية كوهنين ذات خريطة سمة التنظيم الذاتي
Label	تسمية
Learning Rate	معدل التعليم
Logistic Regression	الانحدار المنطقي
Long Short-Term Memory (LSTM)	ذاكرة طويلة قصيرة الأمد
Machine Learning	تعليم الآلة
Many-to-Many	عدة مداخل لعدة مخارج
Many-to-One	عدة مداخل لمخرج واحد
Max Pool	تجميع الأكبر
Maxpooling Layer	طبقة التجميع الأكبر
Motion Energy Image MEI	صورة طاقة الحركة
Motion Feature	سمة الحركة
Motion History Image MHI	صورة تاريخ الحركة
Motion Silhouette Image MSI	صورة هيئة الشخص المتحركة
Multilayer Feedforward Networks	شبكة متعددة الطبقات
Multi-level MHI MMHI	صورة MHI متعددة المستويات
Multi-Streams	تعدد السيول
Normalization	التطبيع
Not Fallen	عدم السقوط

OF-MHI	صورة تاريخ حركة التدفق البصري
One-to-Many	دخل واحد إلى عدة مخارج
One-to-One	دخل واحد إلى مخرج واحد
Optical Flow (OF)	التدفق البصري
Orientation Angle	زاوية الاتجاه
Overfitting	الموائمة الزائدة
Overlapping	تراكب
Person Detection	كشف الشخص
Pooling	تجميع
Projection Histogram	هستوغرام الإسقاط
Random Forest classifier	مصنف الغابة العشوائية
Recall	الاستدعاء
Rectified linear unit (ReLU) function	تابع تنشيط ذو الوحدة الخطية المصححة
Recurrent Neural Networks RNN	شبكة عصبية تكرارية
Regression	انحدار
Reinforcement Learning	التعلم المعزز
Residual network Resnet	الشبكة العصبونية ذات الباقي
RGB	أحمر-أخضر-أزرق
Saliency Maps	خرائط البروز
Self-Feedback	تغذية راجعة ذاتية

Self-Occlusion	التغطية الذاتية
Sensitivity	الحساسية
Short -Time Fourier Transformation STFT	تحويل فورية في الزمن القصير
Sigmoid Function	تابع تنشيط نسبي
Silhouette	هيئة الشخص
Silhouette Image (SI)	صورة هيئة الشخص
Single-Layer Feedforward Networks	شبكات أمامية وحيدة الطبقة
Softmax Function	تابع الاحتمال الأكبر
Specificity	النوعية
Spectrogram	المخطط الطيفي
Stacked Time-Tagged Foreground Image (STTFI)	صورة أمامية مكدسة بعلامات زمنية
Stride	خطوة
Stride History Image SHI	صورة الحركة ذات الخطوة
Supervised Learning	التعلم الخاضع للإشراف
Support Vector Machine SVM	آلة شعاع الدعم
Tanh Function	تابع تنشيط الظل القطعي
Testing Phase	طور الاختبار
timed-Motion History Image tMHI	صورة تاريخ الحركة الموقوتة
Training Phase	طور التدريب
Transfer Learning	نقل التعلم

True Negative TN	سلبی صحیح
True Positive TP	إيجابي صحیح
Unsupervised Learning	التعلم غير الخاضع للإشراف
Validation	الصلاحية
Vanishing Gradient	التدرج المتلاشي
Video Description	وصف الفيديو
View-Based Approach	طريقة تعتمد على العرض

الملخص

ساهم انتشار تقنيات الذكاء الصناعي في العقد الأخير، ورخص الكاميرات وانتشارها الواسع، في الاهتمام بتصنيف حركات وأفعال البشر خاصة تلك الأفعال التي قد يكون لها أثر سلبي في صحتهم مما قد يؤدي في أسوأ الأحوال إلى وفاتهم، وكان لحدث السقوط الحظ الأكبر من الدراسات والبحوث، وخاصة فيما يتعلق بكبار السن، إذ يكون خطر الموت نتيجة السقوط من نصيبهم. لذلك تم في هذا البحث طرح منظومتين جديدتين لكشف السقوط المؤذي.

تعتمد المنظومة الأولى على تطوير الشبكة الهجينة CNN-LSTM، من خلال اقتراح شعاع سمات جديد مشتق من شبكة CNN المدربة يحمل سمات عمومية وتفصيلية، لتمييز ثلاث وضعيات مرحلية (نشاطات جزئية) متتالية متعاقبة تم اقتراحها لتعريف أنواع السقوط المختلفة. يكون هذا الشعاع دخلاً لشبكة LSTM التي تتميز السقوط بعد مرور سلسلة متتالية من ذلك الشعاع، أعطت هذه المنظومة دقة 98.6% وحساسية 100%، متفوقة بذلك على الطرق التي استخدمت شبكات هجينة من النوع CNN-LSTM. تم تدريب هذه المنظومة على مجموعات تدريب عالمية (FDD, URFD, U. Montreal).

تقدم المنظومة الثانية طريقة جديدة دعت صورة أمامية مكدسة بعلامات زمنية Stacked Time-Tagged Foreground Image (STTFI) تحمل سمات زمانية ومكانية تمثل الحركة وكميتها واتجاهها، مثل صور الحركة (MHI, BMI, ITMI)، ولكن لهذه الصور سلبيات، فيه لا تستطيع تمثيل الأجزاء الساكنة من الجسم التي تعتبر نقاط ارتكاز أو الوضعيات الساكنة العابرة أثناء الحركة، إضافة إلى سلبية التغطية الذاتية التي تعاني منها. ولكن صورة STTFI المقترحة تغلبت على هذه السلبيات. استخدمت صورة STTFI في هذه الطريقة كدخل لشبكة CNN قليلة العمق (Alexnet) وعُلمت بطريقة نقل التعلم لتصنف حدث السقوط من عدمه. أعطت هذه المنظومة باستخدام صورة STTFI الجديدة المقترحة دقة 99.02% وحساسية 100%، متفوقة بذلك على الطرق التي استخدمت صور حركة مثل MHI مع شبكة CNN واحدة أو الطرق التي استخدمت

تعدد السيول، أو الطرق التي استخدمت الشبكات الهجينة CNN-LSTM. تم تدريب هذه المنظومة على مجموعات التدريب الثلاثة العالمية إضافة لمجموعة تدريب مصورة لغرض هذا البحث سميت FD22 احتوت على صور لوضعية الإغماء والسقوط من على الأريكة لم تكن موجودة في المجموعات السابقة.

الكلمات المفتاحية: كشف السقوط، كبار السن، صورة تاريخ الحركة MHI، الشبكة العصبونية الالتفافية CNN، التعلم العميق، نقل التعلم، الإبصار الحاسوبي، خلية الذاكرة القصيرة طويلة الأمد LSTM.

المقدمة

حظي تصنيف النشاط البشري في السنوات الأخيرة باهتمام العديد من الباحثين حول العالم، خاصة مع ظهور وانتشار تقنيات الذكاء الاصطناعي التي تشق طريقها بقوة في هذه الأيام، شمل ذلك العديد من النشاطات الحياتية اليومية (Activities of Daily Living (ADLs) التي يقوم بها الإنسان كالمشي [1]، والركض والجلوس والانحناء والاستلقاء وركوب الدراجة [2-4]... إلخ، كما شمل الاهتمام أيضاً الحركات غير الطبيعية التي يؤديها الإنسان بإرادته كحوادث العنف والتسكع والتطفل [5,6]، أو بدون إرادته كحوادث السقوط [5,7-9] التي كان لها النصيب الأكبر، لما لها من أثر بالغ على صحة الأفراد فحسب تقارير منظمة الصحة العالمية [10] يأتي حدث السقوط في المرتبة الثانية كأحد أسباب الوفاة بعد حوادث الطرق، وتتركز النسبة الأكبر بين الأطفال وكبار السن؛ إذ يدفع حب الفضول واستكشاف البيئة المحيطة الأطفال إلى التعرض بكثرة لحوادث السقوط التي إن لم تكن مميتة ربما سببت لهم إعاقة دائمة، أما بالنسبة لكبار السن الذين تجاوزت أعمارهم 60 فإن حوادث السقوط البسيطة قد تكون مميتة لهم، فقد تتسبب بكسر في الحوض أو رضوض وجروح خطيرة تتطلب إقامة في المستشفى، وفي هذه الأيام مع قلة الكوادر الطبية والتمريضية وتوابع جائحة كورونا، سيكون الأمر مرهقاً مما يتطلب مساعدة تقنية حديثة ليستطيع الكادر الطبي القليل العدد أن يقوم بمساعدة كم أكبر من المرضى، كما يمكن للمساعدة التقنية أن تغني كبير السن عن وجود مرافق دائم له في المنزل، خاصة إن كان يعيش بمفرده. لذلك انتشرت في السنوات الأخيرة العديد من الدراسات لتصميم أنظمة تستطيع اكتشاف السقوط بأسرع وقت ممكن - فكلما كانت الفترة التي يبقى فيها الشخص مدداً على الأرض بعد السقوط طويلة زادت احتمال حدوث الوفاة [11] - وإعطاء إشارة إنذار للكادر الطبي ليقوم بالمساعدة أو إجراء اتصال بالإسعاف لتقديم المساعدة لمن يعيشون بمفردهم.

وقد صنفت تلك الطرق تبعاً لنوع الحساسات المستخدمة في عملية كشف السقوط إلى طرق تعتمد الحساسات المحيطية [12,13] وطرق تعتمد الحساسات الملاصقة للجسم [8,3,14-18] وطرق تعتمد على

الإبصار الحاسوبي، والتي ظهرت في السنوات الأخيرة بسبب انخفاض كلفة الكاميرات بشكل عام وانتشارها الواسع أيضاً بين الناس من خلال الأجهزة الخلية.

تقسم طرق الإبصار الحاسوبي من حيث استخلاص السمات الخاصة بالسقوط إلى طرق هندسية وإحصائية [19-27]، وطرق تعتمد على تقنيات التعلم العميق [7,9,28-37] التي زادت في العقد الأخير، أما بالنسبة لمرحلة التصنيف فتقسم إلى طرق تعتمد على العتبات [19, 21-23, 25, 26]، وطرق تعتمد على تعليم الآلة؛ فمنهم من اتبع الطرق التقليدية في تعليم الآلة [20, 24, 27, 37]، ومنهم من اتبع التعلم العميق [7, 9, 28-37].

وعلى اعتبار أن السقوط حركة؛ أي تغير في موضع الشخص تابع للزمن فهو ليس صورة جامدة، إنما تتابع مجموعة من الصور على شكل فيديو يميز تلك الحركة، ورغم ذلك فقد كانت هناك أبحاث تعتمد على الصورة ذاتها في تصنيف السقوط [28-30]، إلا أن تلك الأبحاث أغفلت الجانب الزمني الذي يؤثر في نتيجة التصنيف، بينما اهتمت أبحاث أخرى بالجانب الزمني فقط واستخلاص السمات الزمنية من سرعة وتسارع واتجاه حركة وأشهرها التدفق البصري (Optical Flow (OF [31, 33]، وأغفلت أيضاً أن الشكل الذي يأخذه الشخص أثناء السقوط يؤثر في التصنيف، بينما اهتمت أبحاث أخرى بالصور المتعاقبة لتحديد حدث السقوط [11, 32, 34-36]، سنركز في هذا البحث على بعض الطرق المتبعة في كشف السقوط ليصار إلى اقتراح منظومة كشف سقوط جديدة.

أ. مسوغات البحث وأهدافه

تمت مراجعة العديد من الدراسات التي اهتمت بنشاطات غير طبيعية تحدث للإنسان والتي من أخطرها السقوط، تبين منها أهميته البالغة على صحة الإنسان خاصة كبار السن القاطنين في المنازل بمفردهم أو في دور العجزة أو حتى المرضى في المشافي، فاتخاذ الإجراءات المناسبة بأسرع وقت ممكن، سيخفف حتماً من

أثر ذلك السقوط على صحة الإنسان. يهدف هذا البحث إلى اقتراح طرق جديدة أكثر فعالية في كشف السقوط المؤذي بكاميرات المراقبة ليصار إلى اتخاذ الإجراءات المناسبة.

ب. مناهج البحث وحدوده وأدواته

تم في هذا البحث اتباع أسلوب التعلم العميق لصور كاميرات المراقبة، وذلك لكشف حدث السقوط، ولم يتم التركيز في هذا البحث على الوسائل الإحصائية أو الهندسية في استخلاص السمات، بل كان التركيز منصّباً على طرق جديدة في استخدام التعلم العميق وطرق جديدة للتعبير عن النشاطات البشرية في الحركة والسكون والزمن المرافق لها، وقد استخدم الماتلاب كأداة رئيسية في هذا المجال.

ت. موجز محتويات الفصول القادمة

بعد المقدمة تقع هذه الأطروحة في خمسة فصول:

الفصل الأول: يتم فيه الحديث عن الشبكات العصبونية الاصطناعية بشكل عام، ثم يليها الحديث عن الشبكات العصبونية الالتفافية، ثم شبكات LSTM، ينتهي الفصل بالحديث عن الشبكات الهجينة ومعايير تقييم أداء الشبكات العصبونية.

الفصل الثاني: يتم في هذا الفصل الحديث عن صورة تاريخ الحركة MHI وكل التغيرات والتعديلات التي طرأت عليها، كما يتم الحديث عن محاسنها وعيوبها.

الفصل الثالث: يتم فيه سرد ومناقشة الدراسات التي ناقشت وطرحت موضوع السقوط بإسهاب، كما تمت مناقشة المشاكل والعيوب التي واجهاتها تلك الدراسات والطرق المتبعة.

الفصل الرابع: يُعرض فيه التعديل المقترح على الشبكة الهجينة CNN-LSTM، ثم استخدامها في المنظومة الأولى المقترحة لكشف السقوط المؤذي، التي تم تعليمها من خلال عدة مجموعات تدريب، بعدها تمت مناقشة النتائج ومقارنتها مع الدراسات السابقة.

الفصل الخامس: يُعرض فيه صورة الحركة الجديدة المقترحة STTFI، ثم استخدامها في المنظومة الثانية المقترحة لكشف السقوط المؤذي، التي تم تعليمها من خلال عدة مجموعات تدريب إضافةً لمجموعة مصورة FD22 لغرض البحث، بعدها تمت مناقشة النتائج ومقارنتها مع الدراسات السابقة.

الفصل السادس: الخاتمة ويُقدّم فيها الاستنتاجات والاقتراحات لتطوير البحث.

الفصل الأول

شبكات التعلم العميق

1-1- تمهيد

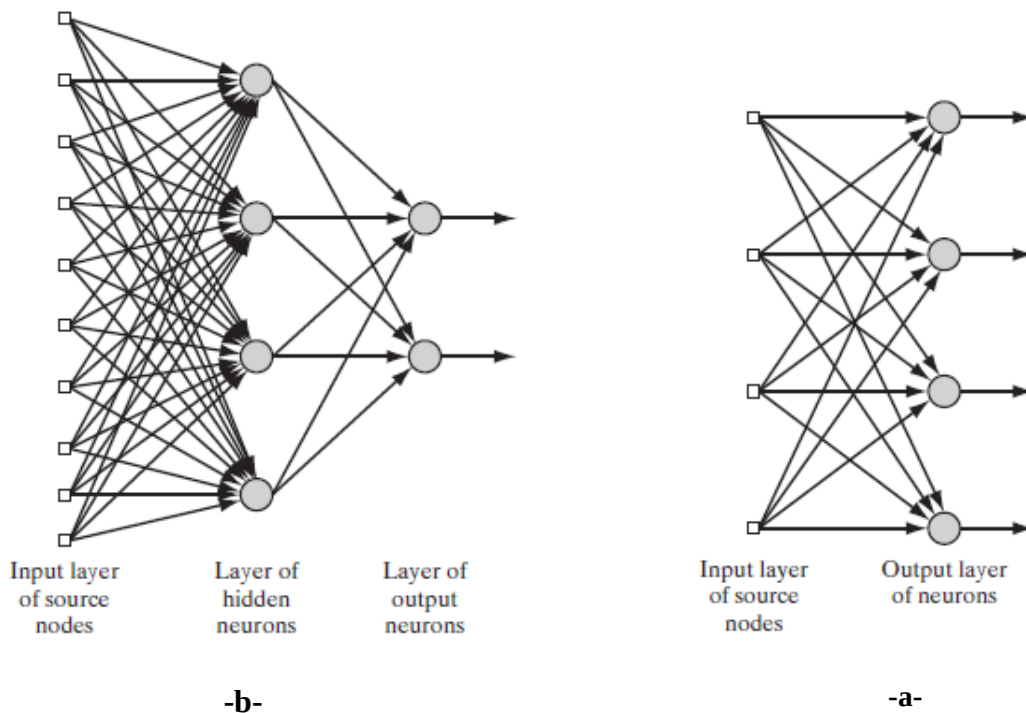
ارتبطت رحلة الإبصار الحاسوبي بفكرة الذكاء الاصطناعي الذي انطلق في منتصف خمسينات القرن الماضي في جامعة دارتموث، حيث تفاعل العلماء الذين أطلقوا ذلك الاسم عليه تفاؤلاً كبيراً [38]، حتى وصل تفاؤلهم إلى أن آلة ذكية ستحل محل الإنسان في غضون بضع سنوات، إلا أن ذلك التفاؤل الهائل اصطدم في تلك الفترة بموارد الحاسوب المتدنية في مواجهة المعضلات المعقدة، مما ولد إحباطاً كبيراً انعكس على الباحثين في هذا المجال، فبدأ الإبصار الحاسوبي وفكرة الذكاء الاصطناعي أيضاً يخوان معاً خلال ستينات وسبعينات القرن الماضي، رغم ظهور بعض الحالات النادرة ولكنها مليئة بالتعقيد الرياضي في مجال الإبصار الحاسوبي في ثمانينات وتسعينات ذلك القرن.

في أواخر التسعينات وبداية القرن الحادي والعشرين تطورت الحواسيب الشخصية تطوراً هائلاً وظهرت شبكة الانترنت مما أتاح ربطاً كبيراً بين جميع أنحاء العالم، ما مكن العلماء من جمع كمٍ كبيرٍ من المعطيات أكثر من أي وقت مضى، وتطورت خوارزميات تعليم الشبكات العصبونية الاصطناعية المقترحة في ثمانينات وتسعينات القرن الماضي، وهذا ما أدى إلى ظهور الاختراق الأول والبذرة الأساس في مجال الإبصار الحاسوبي عام 2012 على يد Krizhevsky وآخرين [39]، فقد اقترح شبكة عصبونية التلافية convolutional neural network (CNN) تدعى Alexnet، إذ كان يعتمد أغلب الإبصار الحاسوبي قبلها على طرق استخلاص سمات إحصائية أو هندسية مقترحة بشرياً مستخلصة من الصورة، ثم استخدام طرق العتبات الرياضية أو طرق التصنيف المختلفة مثل آلة شعاع الدعم (Support Vector Machine SVM)، والانحدار المنطقي (Logistic regression) والغابة العشوائية (Random Forest) والجارك الأقرب (k-nearest neighbor) والشبكات العصبونية الاصطناعية للقيام بالتصنيف، ولكن شبكة Alexnet كما سنرى لاحقاً تقوم باكتشاف سمات الصورة

دون تدخل بشري ومن ثم تقوم بالتصنيف؛ أي أنها تستطيع استخلاص السمات الحرجة للصورة وفهم تعقيدها كما يفعل الدماغ البشري، فليس للعنصر البشري أي دخل إلا إعطاءها مجموعة التدريب Dataset، وقد فتحت هذه الشبكة بفعاليتها في تمييز وتصنيف الصور المجال واسعاً للعديد من الباحثين، فظهرت بعدها العديد من الشبكات العصبونية الالتفافية ذات البنى المختلفة، نذكر منها ZFNET و GoogleNET و VGG و RESNET... إلخ. ولكن قبل الخوض في تلك الشبكة سيتم استعراض بنية الشبكة العصبونية الاصطناعية الأساسية وطرق تعليم الشبكات العصبونية الاصطناعية.

2-1- الشبكات العصبونية الاصطناعية

ظهرت الشبكات العصبونية الاصطناعية في منتصف الخمسينات، حيث ظهرت ثلاثة أنواع لهذه الشبكات، الشبكة وحيد الطبقة Single-Layer Feedforward Networks، والشبكة متعددة الطبقات Multilayer Feedforward Networks، والشبكة العصبونية التكرارية (Recurrent Neural RNN)



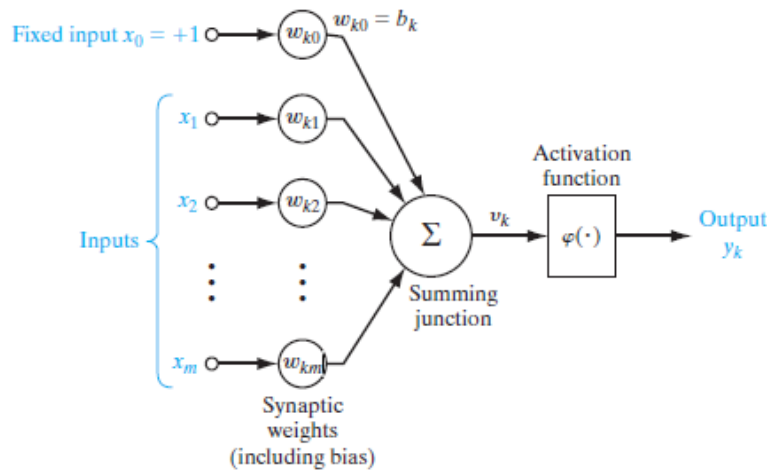
الشكل 1-1 (a) شبكة عصبونية أمامية وحيدة الطبقة، (b) شبكة عصبونية أمامية متعددة الطبقات [41]

(Networks)، تدل كلمة أمامية Feedforward على أن الإشارة تسلك في الشبكة اتجاهاً واحدة من الدخل إلى الخرج فقط [40]. يظهر الشكل (1-1) النوعين الأوليين.

في الشكل (a-1-1) تتكون الشبكة العصبونية البسيطة أو ما تعرف بوحيدة الطبقة من طبقة واحدة هي طبقة الخرج، وعلى الرغم من وجود طبقة الدخل، فإن الطبقات المحسوبة هي تلك التي تحمل تابع تنشيط أما طبقة الدخل ليست كذلك، تابع التنشيط ϕ هو تابع رياضي غير خطي يستخدم لتوليد الخرج اعتماداً على الدخل المعطى له، والدخل المعطى له عبارة عن مجموع المداخل $x_1..x_m$ الموزونة بأوزان w_{km} بالإضافة إلى قيمة الإزاحة b_k ، وتمثل المعادلة (1-1) خرج العصبون في الشبكة العصبونية الاصطناعية

$$y_k = \sum_{j=1}^m w_{kj} * x_j + b_k \quad (1-1)$$

يبين الشكل (2-1) كيفية حساب خرج العصبون.



الشكل 1-2 مخطط يبين كيفية حساب خرج عقدة أو عصبون من الشبكة العصبونية الاصطناعية [41]

أما الشكل (b-1-1) فيبين الشبكة العصبونية متعددة الطبقات: حيث تتألف من طبقة دخل وطبقة وسيطة أولى تدعى بالطبقة الخفية - قد يكون هناك أكثر من طبقة خفية - ثم تليها طبقة الخرج، وتتعدد الطبقات الخفية ويزداد حجم كل طبقة بناءً على نوع المعطيات المراد التعرف عليها حيث تتعرف الطبقات الخفية الأولى على السمات الأساسية بينما تتعرف الطبقات الخفية الأخيرة على السمات الأكثر تعقيداً المكون من السمات المكتشفة في الطبقات السابقة، ويقع على عاتق المصمم اختيار عدد الطبقات وعدد العقد في كل طبقة.

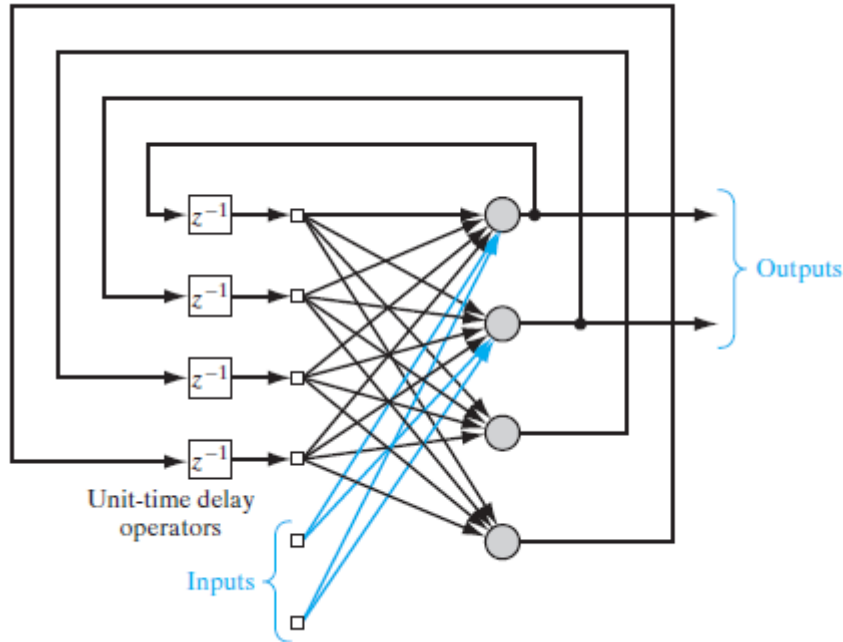
يوصف النوعان السابقان من الشبكات العصبونية بأنهما شبكتان كاملتا الاتصال (Fully connected)،

أي أنه في كل طبقة أي عقدة منها تتصل بجميع العقد في الطبقة التي تليها من خلال الأوزان الخاصة بها.

أما النوع الثالث من الشبكات فهو الشبكات العصبونية التكرارية RNN والتي فيها يغذي الخرج عقدة

معينة أو عدة عقد بناء على تأخير زمني Z^{-1} ، فيكون خرج أي عقدة معتمداً على الخرج السابق والمدخلات

الحالية، وقد يكون ذلك في الشبكات وحيدة الطبقة أو متعددة الطبقات ويبين الشكل (3-1) ذلك.



الشكل 3-1 شبكة تكرارية بطبقة عصبونات خفية [40]

وقد تكون التغذية الراجعة من خرج العقدة للعقدة ذاتها فقط، عندها نسمي تلك التغذية بالتغذية الراجعة

الذاتية self-feedback.

1-3-3-1- تعليم الشبكة العصبونية

تعليم الشبكة هو العملية التي يتم خلالها تغيير قيم الأوزان التي تصل بين الدخل والعصبونات وبين العصبونات في طبقة والعصبونات في الطبقة التي تليها للحصول على الأداء الأفضل الذي يحدده تابع الخطأ أما بالنسبة لتعليم تلك الشبكات فهناك ثلاثة أنواع لتعليم الشبكات العصبونية الاصطناعية:

1-3-3-1- التعلم الخاضع للإشراف Supervised Learning

يتم تدريب الشبكة من خلال تزويدها بالمدخلات وأنماط المخرجات المطابقة لها، وذلك من خلال مجموعات معطيات تدريب Datasets خاصة بالأنماط أو الأصناف المراد تمييزها. وهذا النوع من التعلم هو الأكثر انتشاراً، فمجموعة التدريب تتألف من مدخلات ومخرجات مقابلة لها؛ أي مجموعة معطيات ذات تسميات [41].

يُستخدم هذا النوع من التعلم في حل أحد مسألتين، مسألة التصنيف classification أو مسألة الانحدار Regression بنوعيه الخطي وغير الخطي [42]، حيث تستخدم خوارزمية الانتشار العكسي (Back Propagation)، طرح هذه الخوارزمية Rumelhart وآخرون [43] وقد لاقت انتشاراً واسعاً لأنها تلافت معظم سلبيات الخوارزميات الأخرى، وتطبق هذه الخوارزمية على الشبكة العصبونية الاصطناعية بأنواعها الثلاثة المذكورة سابقاً.

تبدأ هذه الخوارزمية بإعطاء قيم عشوائية للأوزان في الشبكة، ثم تعرض المدخلات من مجموعة التدريب على الشبكة تباعاً، حيث في كل مرة تقوم الشبكة بحساب المخرجات الناتجة عن انتشار المدخلات في الشبكة بدءاً من طبقة الدخل حتى الوصول إلى طبقة الخرج، وهذا ما يعرف بالانتشار الأمامي (forward propagation)، بعدها تقارن قيم المخرجات الناتجة مع القيم الفعلية الصحيحة (Ground Truth) الموجودة ضمن مجموعة التدريب، فينتج عن ذلك إشارة خطأ تُستخدم بواسطة طريقة الانحدار المتدرج (Gradient-descent rule) وقاعدة ديلتا العامة لضبط قيم الأوزان الجديدة ابتداء من الخرج وحتى الوصول إلى الدخل وهذا

هو الانتشار العكسي [44]، يستمر ذلك حتى آخر مُدخل حيث تسمى هذه بدورة التدريب (Epoch)، وتكرر دورة التدريب عدة مرات حتى الحصول على أدنى قيمة للخطأ عندها تتوقف عملية التدريب، إذ يطلق على هذه المرحلة طور التدريب (Training Phase)، تليها مرحلة أخرى هي طور الاختبار (Testing Phase) ويستخدم في هذه المرحلة مجموعة جزئية (مجموعة الاختبار) من مجموعة التدريب لم تستخدم في طور التدريب، وتكرر هذه المجموعة على الشبكة المدربة لتنتج المخرجات المتوقعة والتي تقارن مع المخرجات الحقيقية ضمن مجموعة الاختبار لتعطينا الدقة النهائية للشبكة المدربة [41].

1-3-2- التعلم غير الخاضع للإشراف Unsupervised Learning

التعلم غير الخاضع للإشراف أو التنظيم الذاتي، يتم في هذا النوع من التعلم تدريب الشبكة على إعطاء مخرجات معينة هي تحدد استجابة لمجموعات من المدخلات المتشابهة التي تشكل نمطاً ما غير محدد وعلى الشبكة معرفته؛ بعبارة أخرى، من المفترض أن تكتشف الشبكة السمات البارزة إحصائياً لمجموعة المدخلات، فعلى عكس نموذج التعلم الخاضع للإشراف، لا توجد مجموعة مسبقة من الفئات -الأصناف- التي يتم تصنيف المدخلات إليها؛ فبدلاً من ذلك، يجب أن تطور الشبكة تمثيلها ورؤيتها الخاصة للمدخلات، وفي هذا النوع من التعلم فقط تعطى للشبكة مجموعة المدخلات ولا يوجد مخرجات مقابلة لها، أي أن مجموعة معطيات التدريب بدون تسميات [41].

بالنسبة للتعلم غير الخاضع للإشراف فإن هناك العديد من الخوارزميات للقيام بهذا النوع من التعلم، ولكن فيما يخص الشبكات العصبونية فمن طرق تعليمها خوارزمية (Kohonen Self-Organizing feature Map SOM) [45].

1-3-3- التعلم المعزز Reinforcement Learning

في هذا النوع من التعلم تقوم آلية التعلم المعتمدة ببعض الإجراءات على البيئة من خلال المخارج ومن ثم تحصل على استجابة من تلك البيئة، بناء على هذه الاستجابة التي تبديها البيئة تقوم آلية التعلم بتقييم ما قامت به على أنه جيد (مجزي)، أو سيء (يحتاج عقاباً)، وبناء على ذلك تقوم بتعديل مستويات الشبكة العصبونية ويستمر ذلك حتى الوصول إلى حالة من التوازن لا تحتاج لتعديل تلك الأوزان [41].

1-4- شبكات التعلم العميق

هناك العديد من أنواع شبكات التعلم العميق، لكن تم التركيز في هذا البحث على الشبكات العصبونية الالتفافية CNN وشبكات الذاكرة القصيرة طويلة الأمد LSTM، والشبكة الهجينة CNN-LSTM، لأنها الأكثر استخداماً في تقنيات كشف السقوط.

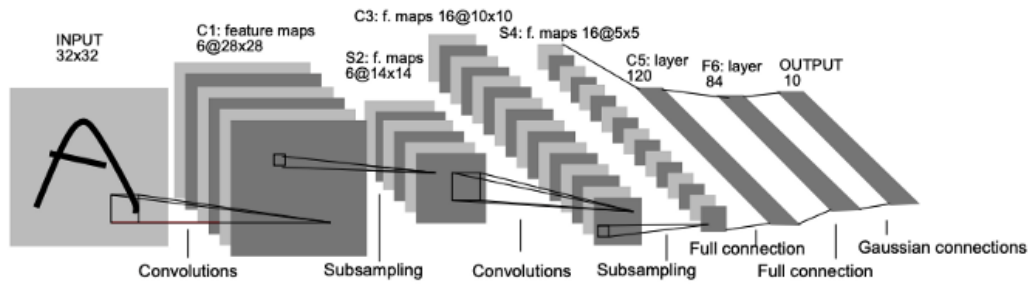
1-4-1- الشبكات العصبونية الالتفافية

استُوحيت فكرة الشبكات العصبونية الالتفافية مما قام به كل من Hubel و Wiesel عام 1959 [46]، فقد اكتشفا أن نظام الرؤية لدى القطط يتألف من نوعين من الخلايا في قشرة الدماغ: خلايا معقدة، وخلايا بسيطة، فالبسيطة تقوم باكتشاف الحواف والخطوط وفق اتجاهات معينة وفي أماكن محددة من الصورة، بينما تقوم المعقدة باكتشاف الأشكال المعقدة مثل الخطوط المنحنية والأشكال الهندسية مهما كان موضعها في الصورة؛ ويعود السبب في ذلك إلى أن الخلية المعقدة تجمع معلوماتها من الخلايا البسيطة التي قبلها، جعل ذلك الاكتشاف Fukushima [47] يبني شبكة عصبونية اصطناعية سماها Neocognitron -عام 1988- تحاكي تلك الخلايا المعقدة والبسيطة، حيث سمى في شبكته طبقة الخلايا المعقدة C، أما طبقة الخلايا البسيطة فسامها S، إذ تقوم الطبقة S باكتشاف السمات البسيطة في الصورة كالحواف والخطوط، ثم تقوم الطبقة C التي تليها

باكتشاف السمات الأعقد، ثم تقوم الطبقة التي تليها باكتشاف سمات أكثر تعقيداً حتى تصل إلى الطبقة الأخيرة التي تحدد نوع الكائن.

LeNet شبكة 1-1-4-1

استفاد من ذلك Lecun وآخرون [48] في أواخر تسعينات القرن الماضي وصمموا شبكة عصبونية اصطناعية لكشف الأحرف المكتوبة بخط اليد، إذ تعتبر تلك الشبكة هي أول الشبكات العصبونية الالتفافية CNN، ويبين الشكل (4-1) تلك الشبكة.



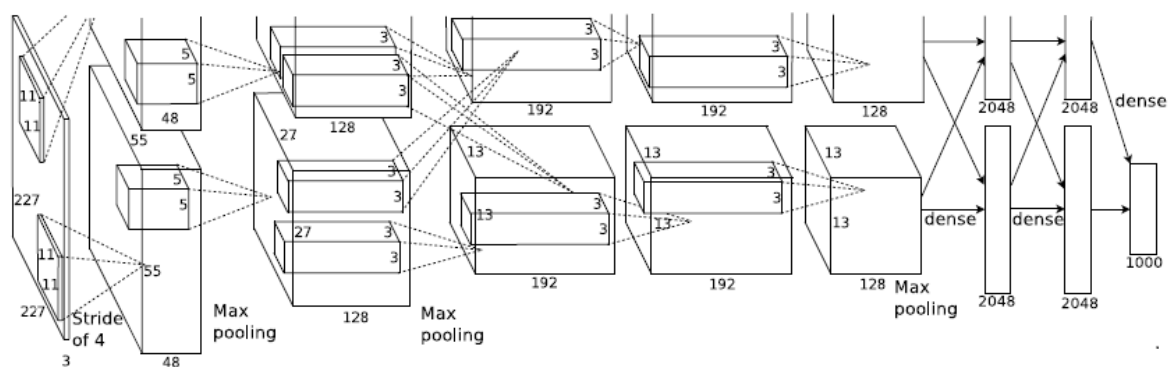
الشكل 4-1 شبكة LeNet العصبونية الالتفافية لتمييز الأحرف [48].

تتألف هذه الشبكة العصبونية كما يبينه الشكل (4-1) من مراحل متتابعة كل مرحلة تحوي مجموعة من الشرائح، وكل شريحة (تسمى بالمرشح) تحوي مجموعة من العصبونات لها أوزان معينة، وبواسطة عملية الطي (convolution) يمر كل مرشح على كامل الصورة، والخرج الناتج يسمى بخريطة السمات (feature map)، إذ تؤمن هذه العملية مقداراً من عدم التأثر بالانزياح أو التشوه أو تغير الأبعاد للشيء الموجود في الصورة والمراد كشف سماته، فمهما كان مكان الشيء في الصورة فإن المرشح سيمر على كل جزء من الصورة ليكتشف فيه السمات الأولية، وبذلك يمكن المحافظة على الموضع الحيزي لمكان تلك السمات المكتشفة؛ بعبارة أخرى يستخدم الطي أو الالتفاف للتعرف على ميزات الصورة بشكل أفضل نظراً لارتباط خرج العصبون بالنقاط المتجاورة في ذلك الجزء من صورة الدخل، وهذا ما يسمح أيضاً بتمثيل أفضل للمناطق المتداخلة.

خرائط السمات في المرحلة السابقة تصبح دخلاً للمرحلة التالية ليُكتشف فيها مجموعة من السمات المركبة من السمات السابقة لها، فنتج خرائط سمات أكثر تعقيداً، وهكذا... حتى المرور على كل المراحل، حيث ينتهي بذلك طور استخراج سمات الصورة تلقائياً دون تدخل من البشر ليأتي بعدها طور التصنيف، إذ تمر خرائط السمات في المرحلة الأخيرة بعملية المد (flatten)، فتتحول بموجبها خرائط السمات الأخيرة إلى طبقة عصبونية واحدة تشكل طبقة الدخل لشبكة عصبونية متعددة الطبقات تقوم بعملية التصنيف.

AlexNet شبكة 2-1-4-1

مرت العديد من السنوات والمحاولات والتحسينات حتى قام Krizhevsky وآخرون [39] بتقديم شبكتهم العصبونية الالتفافية CNN في مسابقة ImageNet Large Scale Visual Recognition Challenge (ILSVRC)، والتي فتحت المجال واسعاً للإبصار الحاسوبي بواسطة الشبكات العصبونية الالتفافية. تتألف هذه الشبكة كما يظهرها الشكل (1-5)، من خمسة طبقات التفافية مخصصة لاستخراج السمات، وثلاثة طبقات عصبونية كاملة الاتصال FC مخصصة لتصنيف الصور إلى ألف صنف.



الشكل 1-5 شبكة AlexNet العصبونية الالتفافية [39].

طبقة الطي الأولى تتألف من مرشحات عددها 96 مرشحاً موزعة على وحدتي معالجة رسوميتين GPU، وأبعاد المرشح الواحد في الطبقة الأولى هي $11 \times 11 \times 3$ ، ودخل هذه الطبقة هو صورة بحجم $227 \times 227 \times 3$ ، وينتج عن خرج هذه الطبقة خرائط سمات عددها 96 وبأبعاد 55×55 تحتوي على السمات

الأولية المستخلصة من الصورة، يطبق على تلك الخرائط تابع تنشيط (active function) وتابع تجميع (pooling function) وتطبيع (normalization)، ثم يمرر الخرج الناتج إلى طبقة الطي الثانية ذات المرشحات 256 وكل مرشح بأبعاد $5 \times 5 \times 96$ حيث ينتج عن هذه الطبقة خرائط سمات عددها 256 وأبعاد كل منها 27×27 ، يطبق على خرج تلك الطبقة أيضاً التوابع السابقة الذكر، التنشيط والتجميع والتطبيع، لتمر إلى كل من الطبقات الثالثة والرابعة والخامسة ولكن بين الطبقتين الأخيرتين لا يطبق تابع التجميع أو التطبيع، والطبقة الثالثة تتألف من 384 مرشحاً بأبعاد $3 \times 3 \times 256$ ، أما الرابعة فتتألف من 384 مرشحاً بأبعاد $3 \times 3 \times 384$ ، وأخيراً الخامسة المؤلفة من 256 مرشحاً بأبعاد $3 \times 3 \times 384$ ، وخرجها يطبق عليه التجميع، ويأتي بعدها دور عملية المد لتتكون بموجبه الطبقة السادسة المؤلفة من 4096 سمة تفصيلية تكون دخلاً للطبقة السابعة كاملة الاتصال FC7 تليها الثامنة التعميمية FC8 يليها تابع (softmax) لإعطاء قيمة احتمالية لكل من الأصناف الألف في الطبقة الأخيرة.

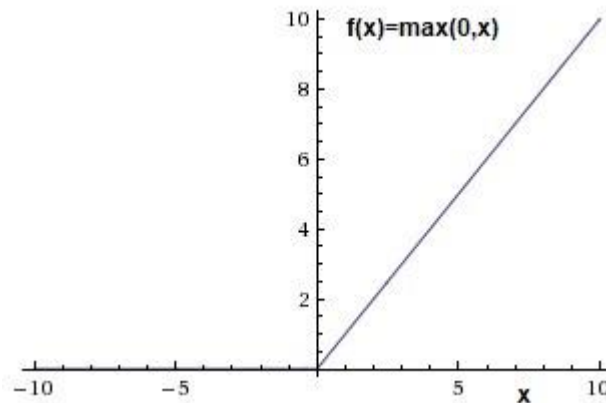
تختلف شبكة Alexnet عن شبكة LeNet في استخدام ثلاث طبقات التقافية (الثالثة والرابعة والخامسة) بدون طبقة تجميع بينها، إضافة إلى استخدام تابع التنشيط ReLU على مخارج جميع العصبونات في الشبكة، كما تبين لمصمم الشبكة فيما بعد بأن طبقة التطبيع ليست بذات فائدة.

هناك العديد من الشبكات العصبونية الالتفافية التي ظهرت بعد شبكة Alexnet مثل VGG و ResNet و InceptionV3 و Xception وجميعها أعمق من شبكة AlexNet وأعقد في البنية، وبسبب بساطة وسرعة تدريب Alexnet فقد اختيرت في هذه الدراسة.

يلاحظ من تلك الشبكات تكرار مجموعة من الطبقات بترتيب معين، لذلك سيتم شرح كل طبقة بشيء من التفصيل.

1-4-2- طبقة الالتفاف أو الطي Convolution

هي الطبقة الرئيسية والأساسية التي سميت بناءً عليها الشبكات العصبونية الالتفافية، وتحتوي كل طبقة كما أسلف ذكره على مجموعة من المرشحات، كل مرشح عبارة عن مصفوفة مربعة -أصغر مساحة من الصورة- يحمل كل عنصر فيها قيمة محددة تسمى الوزن، وبعملية الطي بين المرشح وجزء مقابل له من الصورة ينتج خرج العصبون الذي يمرر على تابع تنشيط لينتج الخرج النهائي، وجميع الشبكات العصبونية الالتفافية تستخدم تابع النشيط ReLU والذي يمثله الشكل (1-6). تتميز هذه الطريقة عن الطبقة كاملة الاتصال بأن مجموعة الأوزان المتمثلة بأحد المرشحات تمر على جميع أجزاء الصورة، مما يمكنها من كشف ذات السمات إن وجدت بعدة أماكن في الصورة، على عكس الطبقة كاملة الاتصال التي تحتاج كم هائل من الأوزان لتصل كل عصبونات الدخل بالعصبونات التي ستنتج الخرج [49].



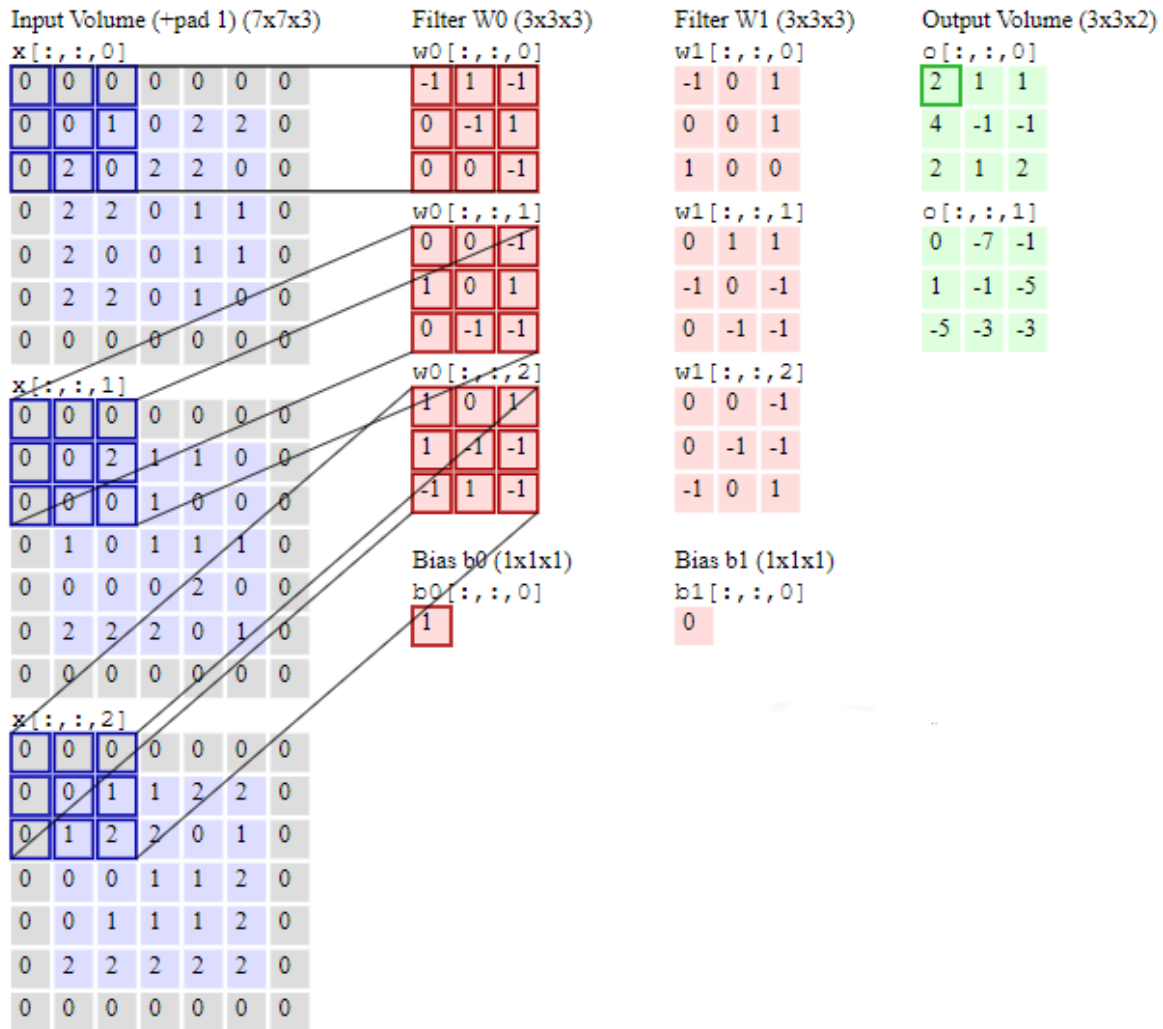
الشكل 1-6 تابع التنشيط ReLU المستخدم في الشبكات العصبونية الالتفافية [50].

يمكن توضيح عملية الطي أو الالتفاف من خلال الشكل (1-7)، وفيه الدخل عبارة عن صورة بأبعاد $5 \times 5 \times 3$ مع حشو P بمقدار 1 وبقية عناصر حشو صفرية، أما طبقة الطي فمؤلفة من مرشحين W_0 بأبعاد $3 \times 3 \times 3$ و W_1 بأبعاد $3 \times 3 \times 3$ ، خرج المرشح W_0 هو O_0 وخرج المرشح W_1 هو O_1 أي يتم الحصول على خريطتي سمات، يكون كل خرج هو حاصل تطبيق المعادلة (1-1) السالفة الذكر، يتم عمل طي بين المرشح والدخل بخطوة (Stride) مقدارها 2، بحيث يمر المرشح على كل جزء موافق له من صورة الدخل ويتم ضرب

عناصر المرشح بمقابلاتها في الدخل وجمع النتيجة وإضافة الإزاحة ثم تطبيق التابع ReLU فينتج خرج الطبقة الجديد، ونلاحظ بأن طبقة الدخل والمرشح لهما ذات العمق، بينما الخرج الجديد عمقه يساوي عدد مرشحات الطبقة أي أن الخرج الجديد O سيكون بأبعاد $3 \times 3 \times 2$. ولمعرفة أبعاد الخرج الجديدة، الارتفاع H والعرض W والعمق D يمكن تطبيق المعادلة (2-1)

$$H \times W \times D = \left(\frac{(N-F+2P)}{S} + 1 \right) \times \left(\frac{(N-F+2P)}{S} + 1 \right) \times K \quad (2-1)$$

حيث تمثل K عدد المرشحات في تلك الطبقة، F بعد المرشح، N بعد الصورة أو الدخل، S الخطوة المستخدمة في الطي، و P عدد بكسلات الحشو في الجانب الواحد.



الشكل 1-7 رسم توضيحي يبين كيفية حساب خرج طبقة الطي في الشبكات العصبونية الالتفافية [49].

1-4-3- طبقة التجميع pooling

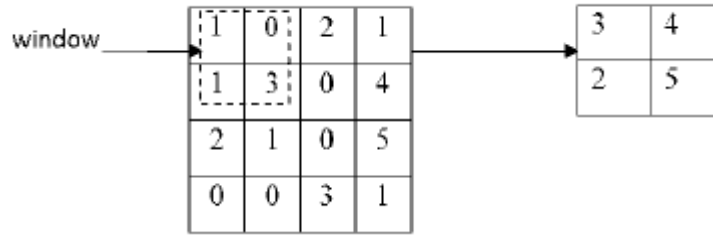
يلاحظ من الشبكات السابقة استخدام طبقة تدعى طبقة التجميع، تأتي بعد طبقة الطي، لا تحوي طبقة التجميع على أي أوزان أو متوسطات، إنما مهمتها هي تخفيض أبعاد الخرج، مع المحافظة على السمات الأساسية الأقوى، وبذلك تزيد من سرعة التعلم بسبب انخفاض عدد المتوسطات المراد تعلمها، كما تتخلص من سيئة الموائمة الزائدة overfitting، وتزيد من خاصية التعميم generalization التي تحتاجها معظم الشبكات العصبونية الاصطناعية، إذ تكمن مهمة المرشح التجميعي في انتقاء قيمة معينة من بين القيم الموجودة ضمن نطاق

المرشح عندما يمر على كل بقعة من خرج الطبقة التي يعمل عليها، فأوزان عناصر المرشح جميعها بقيمة 1. هناك عدة أنواع من المرشحات التجميعية التي يمكن رؤيتها في الشبكات العصبونية الالتفافية.

1-3-4-1- Maxpooling مرشح التجميع

يقوم هذا المرشح بأخذ القيمة الأكبر من بين القيم التي تقع ضمن نطاق المرشح عندما يمر على كل بقعة من خرج الطبقة التي يعمل عليها، فبعد عملية الطي أو الالتفاف لا يعود هناك أهمية بالغة للموضع الدقيق للسمة المكتشفة ضمن الصورة لذلك يمكن تخفيض الأبعاد وانتخاب القيمة الأعلى ضمن بقعة معينة لتمثل السمة المكتشفة [51]، ويبين الشكل (1-8) عملية التجميع بالقيمة الأعلى.

يستخدم هذا المرشح في شبكة AlexNet، إذ تكون طبقة التجميع مؤلفة من مرشح تراكبي (overlapping) بأبعاد 3×3 بخطوة 2 حيث بموجبها تنخفض أبعاد الخرج إلى النصف، وقد أعطى استخدام هذا المرشح نتائج أفضل من ناحية معدل الخطأ top-1 بتحسّن قدره 0.4%، و0.3% من أجل معدل الخطأ top-5، مقارنة مع النتائج الحاصلة عندما استخدم مرشح 2×2 بخطوة 2؛ أي بدون تراكب [39]. ويستخدم هذا المرشح في العديد من الشبكات العصبونية الالتفافية (AlexNet, VGGNet, ResNet, InceptionV3, Xception) [49].



الشكل 1-8 رسم توضيحي يبين كيف يحسب خرج طبقة Maxpooling [51].

1-4-3-2- مرشح التجميع Average Pooling

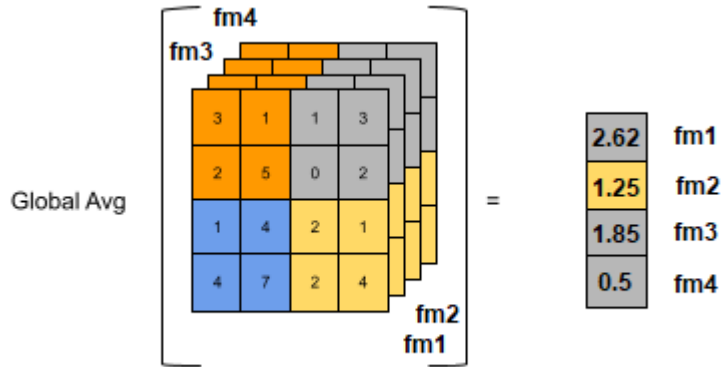
يعمل هذا المرشح على أخذ القيمة الوسطية لقيم البكسلات الواقعة في نطاقه عندما يمر على كل بقعة من خرج الطبقة التي يعمل عليها، يستخدم هذا المرشح في العديد من الشبكات العصبونية الالتفافية مثل (Resnet, InceptionV3).

1-4-3-3- مرشح التجميع Global average pooling

يستخدم هذا النوع من المرشحات بدل مرحلة المد، حيث خرائط السمات الأخيرة الناتجة عن طبقة الالتفاف الأخيرة تخضع لهذه العملية من الترشيح والتي بموجبها ينتج شعاع سمات جديد فيه كل عنصر أو سمة عبارة عن المتوسط الحسابي لجميع القيم المقابلة في أحد خرائط السمات تلك، فلو كان ناتج طبقة الطي الأخيرة هو خرج $4 \times 4 \times 4$ ؛ أي أربعة خرائط سمات (fm1..fm4) كل واحدة بأبعاد 4×4 ، فإن ناتج طبقة الترشيح التجميعي ستكون شعاعاً مؤلفاً من أربعة عناصر كل عنصر يعبر عن المتوسط الحسابي للسمات في خريطة من الخرائط الأربعة، ويوضح الشكل (1-9) هذه العملية.

تستخدم هذه الطريقة بدلاً من مرحلة المد التي تربط بين خرج طبقة الطي الأخيرة والطبقات العصبونية كاملة الاتصال المستخدمة للتصنيف، حيث باستخدامها يتم التخفيف من الأوزان المراد تعلمها والمرتبطة بطبقة المد، إذ يمكن أن يرتبط خرج طبقة المرشح هذه مباشرة بطبقة softmax بدون الحاجة إلى أي من الطبقات العصبونية كاملة الاتصال، وهذا ما يساعد في التخفيف من الموائمة الزائدة التي تعاني منها الطبقات كاملة

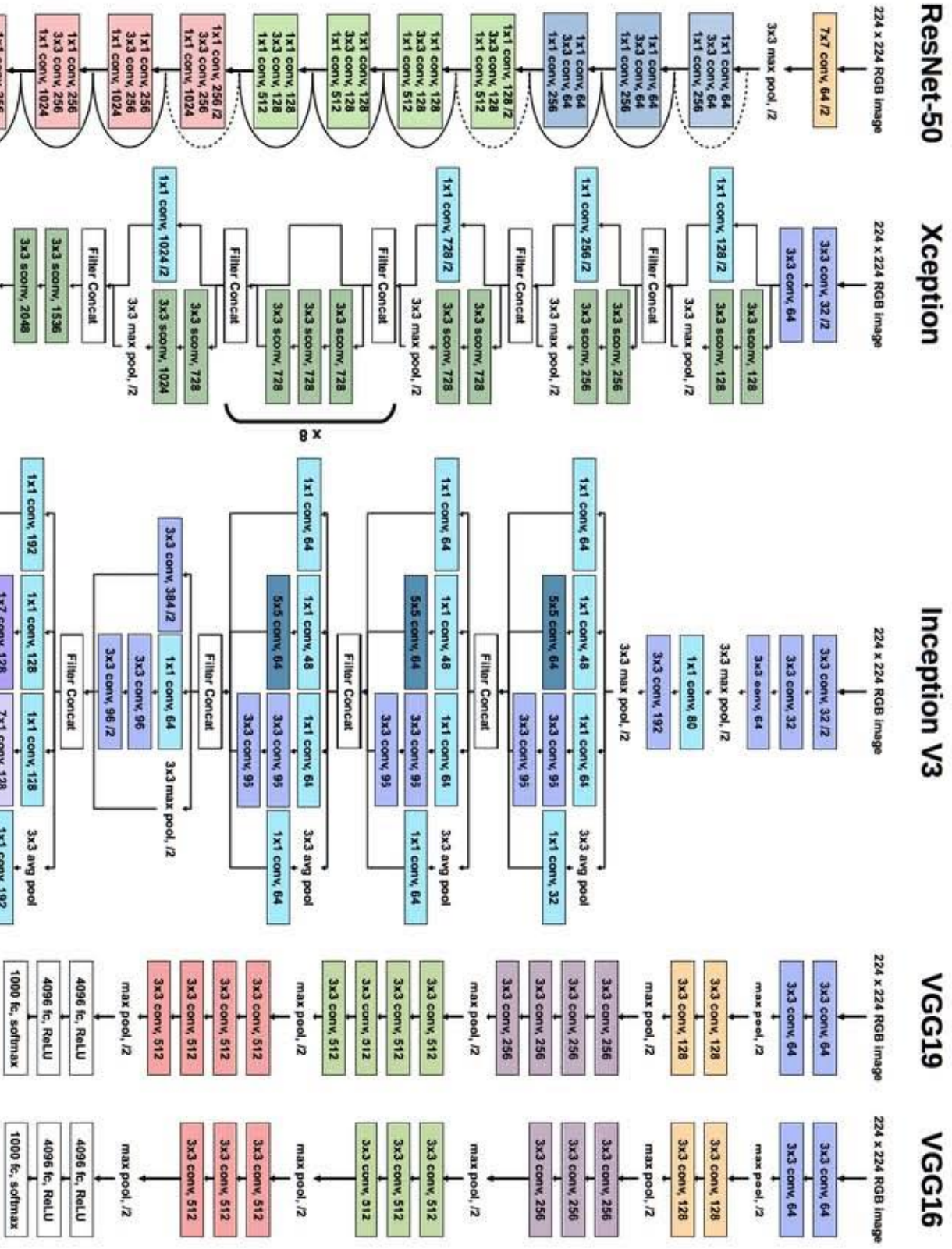
الاتصال والتي كان يتم التخفيف منها سابقاً باستخدام الاسقاط العشوائي (Dropout) لبعض الأوزان أثناء مرحلة التعلم، كما يؤمن استخدام طبقة الترشيح هذه التوافق بين طبقات الطي والأصناف المراد تمييزها [52].



الشكل 1—9 يوضح ناتج تطبيق المرشح التجميعي ذو المتوسط العام على خرائط السمات [52].

ويستخدم هذا المرشح أو طبقة الترشيح هذه في الشبكات العصبونية الالتفافية InceptionV3 وXception.

يبين الشكل (10-1) بنية بقية الشبكات العصبونية الالتفافية. يلاحظ من تلك الشبكات العمق الكبير والتعقيد في التصميم، وهذا ما ينعكس على عدد المتوسطات المراد تعلمها، والمدة الزمنية الطويلة لمرحلة التعليم، لذلك تم اختيار شبكة Alexnet لتكون في المنظومة المراد اقتراحها، فهي مقارنة بتلك الشبكات أقل عمقاً وأبسط تركيباً، ويبدو من الشكل (10-1) بأن شبكة VGG-16 هي الأقل عمقاً بعد شبكة Alexnet حيث اقترحت في عام 2014 ثم تلتها شبكة VGG-19 كذلك ظهرت شبكة Inception في ذات العام ولكنها أكثر عمقاً وأقل متوسطاتٍ، أما في عام 2015 فقد ظهرت شبكة ResNet بأنواعها المختلفة وكانت عميقة أيضاً، كذلك في عام 2017 ظهرت شبكة Xception، إن جميع الشبكات السابقة عدا شبكة VGG تخلصت من الطبقات كاملة الاتصال -كما أسلف ذكره- من خلال طبقة التجميع ذات المتوسط العام [53].



الشكل 10—1 بنى عدة شبكات التفاضلية مختلفة [53] .

1-4-4- نقل التعلم Transfer Learning

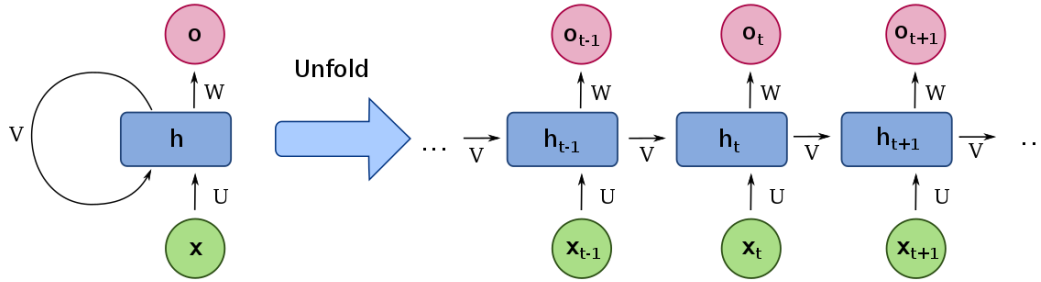
هو أسلوب لنقل قيم المتوسطات المعلمة في شبكة عصبونية التلافية ما على مجموعة بيانات معينة لأداء مهمة معينة إلى قيم جديدة معدلة لتلك المتوسطات لمجموعة بيانات مغايرة للأولى وبمهمة أخرى، يستفاد من هذه الطريقة مع الشبكات العصبونية التلافية المدربة مسبقاً على مجموعة كبيرة من البيانات إذا ما أريد الاستفادة مما تعلمته الشبكة من سمات لتعديلها بحيث تتعلم على مجموعة بيانات جديدة أصغر متعلقة بمهمة جديدة، فتتعلم في وقت أقصر من ذلك الوقت اللازم لتعليمها من البداية بقيم متوسطات عشوائية [54].

إن هذا المفهوم في التعلم انطلق من طبيعة الشبكات العصبونية التلافية، فقد لوحظ بأن تلك الشبكات عند تعليمها تقوم طبقات الطي الأولى بتعلم سمات وخواص عامة للصور لا ترتبط بشكل ضيق بمجموعة الصور المدربة عليها، إنما يتم التخصيص وتعلم السمات الخاصة بتلك المجموعة من الصور في الطبقات الأخيرة من الشبكة العصبونية التلافية [53]؛ بمعنى آخر تحوي طبقات الطي الأولى متوسطات تميز الخواص العمومية للصور من حواف وزوايا وخطوط أفقية وعمودية ومائلة، نهاية بالطبقات الأخيرة التي تميز سمات خاصة بمجموعة صور التدريب الخاصة بغرض معين كرأس شيء ما أو طرف من أطرافه.

1-5- شبكات الذاكرة القصيرة طويلة الأمد (Long Short-Term Memory LSTM)

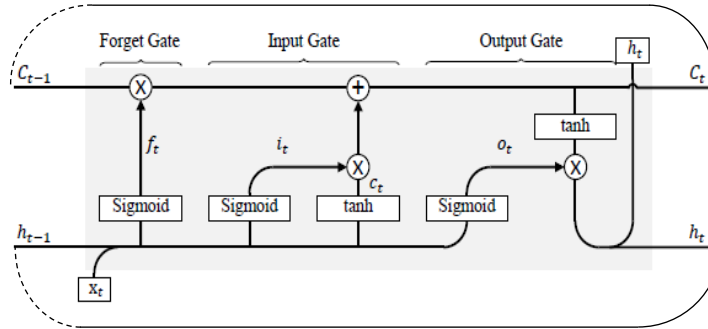
تعتبر الشبكات العصبونية التكرارية RNN التي ذكرت آنفاً أساساً لهذا النوع من الشبكات، فهي تتضمن حلقات تغذية خلفية لإعادة استعمال الأوزان، ولها ذاكرة حالة، وكما أن شبكات CNN تشارك الأوزان ضمن المجال الحيزي من خلال المرشحات وعملية الطي، فإن شبكات RNN تشارك الأوزان ضمن المجال الزمني من خلال التغذية الخلفية؛ أي تُحسب الأوزان في مرحلة معالجة معينة، ثم تُستعمل هذه الأوزان الناتجة ضمن مرحلة المعالجة القادمة. وهكذا في الأوزان الناتجة في المراحل التالية. إن مشاركة الأوزان خلال عدة مراحل يتيح لهذه الشبكات معالجة وتمثيل الأصناف بكفاءة في مجموعة المعطيات المتسلسلة [55]، يبين الشكل (1-11) بنية

شبكة RNN الأساسية، وعملها بمرور الزمن، إذ يلاحظ أنها تتألف من مرحلة دخل x ومجموعة طبقات خفية h ذات تغذية راجعة، ثم تليها طبقة الخرج o .



الشكل 1-11 بنية الشبكة RNN الأساسية وحالتها أثناء مرور الزمن [57]

تعاني هذه الشبكة أثناء مرحلة التعلم - باستخدام طريقة الانتشار العكسي خلال الزمن BPTT والانحدار المتدرج مع وجود سلاسل دخل طويلة - من مشكلتين تؤديان في حال حدوثهما إلى فشل عملية التعلم؛ الأولى وهي التدرج المنفجر (Exploding gradient) حيث أن قيمة التدرج تصبح أكبر من الواحد مع مرور الزمن، مما يؤدي إلى عدم تعلم الأوزان وحدث حالة عدم استقرار في حسابها، تم حل هذه المشكلة من خلال عملية اقتطاع التدرج (Gradient clipping)، أما المشكلة الثانية فهي التدرج المتلاشي (Vanishing gradient)، فعلى عكس السابق يتلاشى التدرج مع مرور الزمن عندما تكون قيمته أصغر من الواحد، ليصبح في النهاية صفراً، وعندها تقشل عملية التعلم لعدم قدرة خوارزمية BPTT على تحديث الأوزان في الشبكة، تم حل هذه المشكلة من خلال اقتراح البنية الشبكية LSTM عام 1997 [56]، إذ تستطيع الاحتفاظ بالسياق الطويل للسلسلة المتتابعة للحالات السابقة بحيث تعتمد عليه في التنبؤ بالخرج الجديد، كما تستطيع حذف السياق غير المهم في تلك السلسلة المتتابعة للحالات السابقة، والذي لا يكون ذو ارتباط وثيق ولا يفيد في التنبؤ الجديد بالخرج التالي، تتكون خلية LSTM من عدة شبكات عصبونية كل شبكة مسؤولة عن مهمة معينة في الخلية، ويبين الشكل (1-12) خلية LSTM.



الشكل 12-1 خلية LSTM العصبونية التكرارية [57]

نرى من الشكل (12-1) أن هذه الخلية تتألف من ثلاث مراحل: مرحلة الدخل، ومرحلة النسيان، ومرحلة الخرج، إذ أضيفت مرحلة النسيان إلى شبكة RNN وعُدلت مرحلة الدخل فيها بحيث تشكلان الخلية (Cell) C ، وهذه الشبكة الجديدة أطلق عليها شبكة LSTM أو خلية LSTM، تقيد الخلية في التخلص من مشكلة RNN في تعلم السلاسل الطويلة، حيث لا تستطيع RNN إلا تعلم المعلومات المرتبطة بضع خطوات إلى الوراء، بينما إضافة الخلية C أعطتها قدرة على حفظ معلومات أطول، لذلك سميت هذه الشبكة بشبكة الذاكرة القصيرة الطويلة الأمد، أما المراحل التي ذكرناها فهي لضمان العمل الأمثل للخلية، فمرحلة الدخل تأخذ شعاع الطبقة الخفية المعبرة عن الحالة السابقة للشبكة h_{t-1} مع الدخل الحالي x_t لتشكل شعاع دخل إلى شبكة عصبونية ذات الأوزان W_c حيث يكون خرجها هو شعاع خلية الذاكرة c_t في الخطوة t ، ولكن ذلك قد يؤدي إلى إشباع الخرج مع تعاقب تكرار الشبكة، لذلك استخدم تابع $\tanh$ لجعل الخرج قياسياً دائماً في المجال $[-1, 1]$ ، وهذا الخرج c_t قد لا يكون هناك حاجة لحفظه كله في خلية الذاكرة لذلك يضرب بتابع البوابة i_t المستنتج أيضاً من شعاع $[h_{t-1}, x_t]$ المضروب بأوزان شبكة عصبونية W_i متبوعة بتابع sigmoid بحيث تكون قيم عناصر الشعاع i_t محصورة بين 0 و 1 فتعمل كالبوابة عند ضربها بشعاع الخرج c_t ، وهكذا يمر منه ما يراد حفظه ليستخدم في الخطوة التالية مضافاً إلى ما سيمر من حالة الذاكرة السابقة C_{t-1} من خلال مرحلة النسيان التي تتألف من الشعاع $[h_{t-1}, x_t]$ الذي يضرب بأوزان شبكة عصبونية W_f خرجها متصل بتابع sigmoid ليكون شعاع بوابة

النسيان f_t الذي سيحدد ما سيمر من الحالة السابقة للذاكرة C_{t-1} وذلك من خلال ضربها به، وبالنهاية يشكل حاصل جمع $i_t \odot C_t$ مع $f_t \odot C_{t-1}$ شعاع الذاكرة النهائي C_t والذي يمر على المرحلة الأخيرة وهي مرحلة الخرج، تتألف هذه المرحلة من شعاع البوابة o_t الناتج من شبكة عصبونية بأوزان w_o مضروبة بالشعاع $[h_{t-1}, x_t]$ يتلوها تابع sigmoid، إذ سيتحكم هذا الشعاع بما يراد أن يظهر كخرج للخلية من شعاع الذاكرة الناتج. الخط المنقطع في الشكل (12-1) الذي يصل C_t بـ C_{t-1} هو التأخير الزمني بمقدار خطوة بحيث يكون C_t هو شعاع الذاكرة كدخل في الخطوة التالية $t+1$ ، وذات الشيء ينطبق على شعاع الحالة h_t .

أما المعادلات المعبرة عن خلية LSTM فهي:

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f) \quad (3-1)$$

$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i) \quad (4-1)$$

$$o_t = \sigma(w_o \cdot [h_{t-1}, x_t] + b_o) \quad (5-1)$$

$$c_t = \tanh(w_c \cdot [h_{t-1}, x_t] + b_c) \quad (6-1)$$

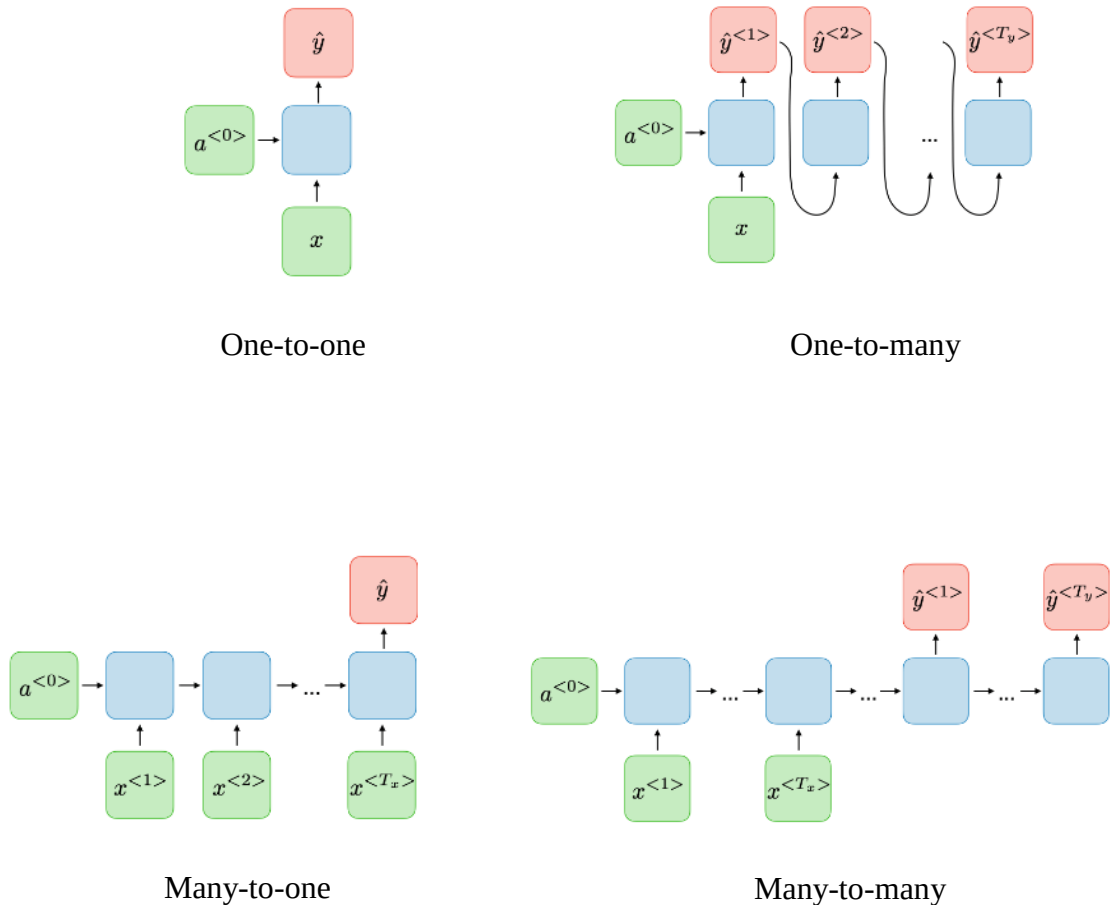
$$C_t = f_t \odot C_{t-1} + i_t \odot c_t \quad (7-1)$$

$$h_t = o_t \odot \tanh(C_t) \quad (8-1)$$

حيث b_f, b_i, b_o, b_c هي أشعة الانحياز ضمن الشبكات التي ذكرت آنفاً.

تجدر الإشارة إلى أن الشبكات العصبونية التكرارية تتنوع تبعاً لعدد المداخل والمخارج، فمنها من له دخل واحد وخرج واحد فيطلق عليها واحد-إلى-واحد (One-to-One)، وهي الشبكات التقليدية المعروفة، ومنها من له دخل واحد وعدة مخارج (One-to-Many) كمولد الموسيقى، ومنها من له عدة مداخل ومخرج

وحيد (Many-to-One) كمصنف المشاعر، ومنها من له عدة مداخل وعدة مخارج (Many-to-Many) كالمترجمات من لغة إلى أخرى. يبين الشكل (1-13) هذه البنى المختلفة [57].

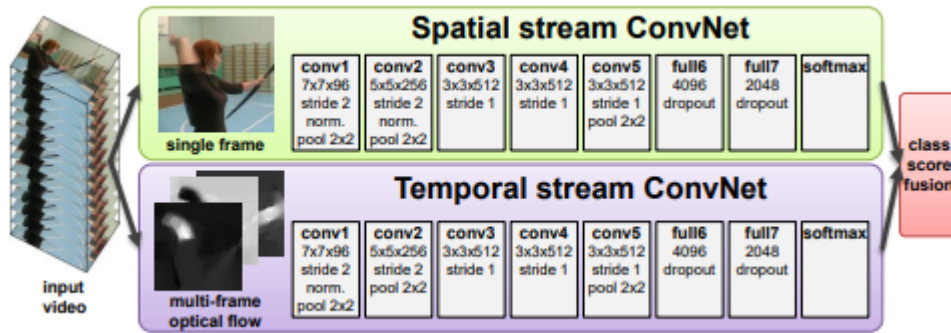


الشكل 1-13 البنى المختلفة لشبكات RNN [57].

1-6- الشبكات الهجينة CNN-LSTM

كما ذكر آنفاً، فإن شبكة CNN تستطيع اكتشاف السمات الحيزية المكانية الموجودة في الصورة، بينما لا تستطيع اكتشاف السمات الزمنية التي تربط الإطارات المتعاقبة، فهي تصلح بشكل كبير لتصنيف الصور أو اكتشاف الكائنات والأشياء ضمن الصور، ولكنها لا تستطيع اكتشاف سمات الربط الزمني بين الإطارات في

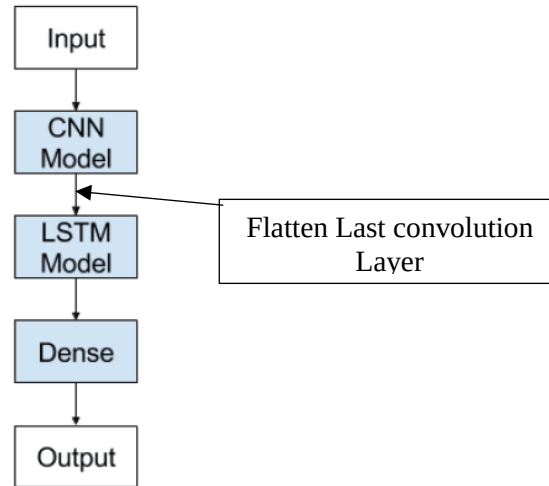
الفيديو التي تكون هامة لتحديد الأفعال التي تقوم بها الأشياء المكتشفة ضمن الصور المتعاقبة، لحل هذه المشكلة اقترح Symon وآخرون [58] في عام 2014 شبكة لتمييز الأفعال البشرية تتألف من جزأين: الأول هو شبكة CNN تأخذ الصور من الفيديو لتكشف السمات المكانية، أما الثاني فهو شبكة CNN لكن دخلها عبارة عن صور التدفق البصري الذي يكشف السمات الزمانية. يتم دمج خرجي الشبكتين في شبكة عصبية عميقة لإعطاء التصنيف النهائي للفعل البشري، كان هذا بداية لشبكات تعدد السيول multi-stream، حيث يبين الشكل (1-14) هذه الشبكة.



الشكل 1-14 شبكة ثنائية السيول لتصنيف الفيديو [58].

إلا أن تلك الشبكة لا تستطيع تصنيف الأفعال البشرية التي تتطلب إطارات طويلة - زمنياً طويلاً- مؤلفة من العديد من الأفعال البشرية الصغيرة، مثل حدث عيد الميلاد، فهو يمتد على العديد من الإطارات ويتضمن العديد من الأفعال الصغيرة (قول الأمنية، إطفاء الشموع، أكل الكعكة) التي بمجمعتها تؤدي إلى التصنيف النهائي "عيد ميلاد" [59]، وبما أن شبكة LSTM تستطيع اكتشاف السمات الزمنية التي تربط المداخل المتسلسلة على المدى الطويل، فقد استخدمت في نوع جديد من الشبكات يدعى الشبكات الهجينة CNN-LSTM التي تتألف من شبكة CNN تقوم باكتشاف السمات المكانية في الصور، تليها شبكة LSTM لكشف السمات الزمانية التي تربط بين الأطر المتعاقبة، يتم الوصل بين الشبكتين بأخذ خرج آخر طبقة - تحمل سمات عمومية - قبل طبقة التصنيف في شبكة CNN وجعله شعاعاً (مده) ليصبح كمدخل لشبكة LSTM التي تليها، بهذه الطريقة يتم كشف السمات الزمانية والمكانية في الفيديو، بحيث يمكن من خلال تلك الشبكة تصنيف الأفعال

البشرية، لكن ذلك الشعاع يحمل سمات عمومية فقط ويهمل السمات التفصيلية التي تصف الغرض بشكل أدق، ويبين الشكل (15-1) هذه الشبكة.



الشكل 15—1 بنية الشبكة الهجينة CNN-LSTM [60].

لقد ظهرت العديد من الشبكات الهجينة من هذا النوع في السنوات الأخيرة وبتشكيلات مختلفة تبعاً للغرض المراد من الشبكة القيام به، فمنها ما اهتم بتصنيف النشاط (Activity Recognition) [59]، ومنها ما اقترح بنى متنوعة لكل من تصنيف النشاط وشرح الصورة (Image captioning) ووصف الفيديو (Video description) [61]، ومنها ما تنبأ بالحالة الجوية [62]، ومنها ما اهتم بالتنبؤ باستهلاك الطاقة الكهربائية [63, 64]، ومنها ما تنبأ بالأورام السرطانية الدماغية [65]، ومنها ما استخدم للتنبؤ بتوسع المدن من خلال صور الأقمار الصناعية [66].

7-1- معايير تقييم عملية تعليم الشبكة

يتم تقييم أداء الشبكة العصبونية الاصطناعية من خلال ثلاثة معايير أداء رئيسية وهي الدقة (Accuracy) والحساسية أو الاستدعاء (Sensitivity أو Recall) والنوعية (Specificity)، تعطى بالعلاقات التالية [67]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (9 - 1)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (10 - 1)$$

$$Specificity = \frac{TN}{TN + FP} \quad (11 - 1)$$

حيث TP (True Positive) تعني عدد مرات التي كان فيها التصنيف المتوقع صحيحاً، و TN (True Negative) عدد المرات التي كانت فيها التصنيفات الأخرى المتوقعة صحيحة أيضاً، أما FP (False Positive) هو عدد المرات التي يتم فيها تصنيف الصنف المرغوب بذاته بينما حقيقة هو من الأصناف الأخرى، وأخيراً FN (False Negative) وهو عدد المرات التي يتم التصنيف على أنها ليست الصنف المرغوب بينما هي في الحقيقة الصنف المرغوب.

تقيس الدقة قدرة الطريقة أو الشبكة على تمييز أي من الأصناف المرغوبة بشكل صحيح، بينما تحدد الحساسية قدرة الطريقة على تمييز الصنف المرغوب بشكل صحيح عن الأصناف الأخرى، أما النوعية فهي تحدد قدرة تلك الطريقة على تمييز الأصناف الأخرى بخلاف الصنف المرغوب بشكل صحيح.

يمكن جعل كل من TP, TN, FP, FN مشمولة في جدول واحد يدعى مصفوفة الالتباس (confusion Matrix)، حيث تمثل التسميات على الأسطر الأصناف الحقيقية (True ground)، بينما تسميات الأعمدة فتمثل الأصناف المتوقعة، أما المربعات فمحتواها يمثل قيم كل من التسميات الأربعة السابقة TP, TN, FP, FN. ويبين الشكل (16-1) هذا الجدول.

		Predicted	
		Positive	Negative
Ground-Truth	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

الشكل 1—16 رسم يبين مصفوفة الالتباس [67]

1-8- الخلاصة

لقد تم الحديث في هذا الفصل عن شبكات التعلم العميق، إذ تم شرح الشبكة العصبونية العميقة منذ انطلاقتها وتعليمها التصنيف لأصناف ذات سمات مستخرجة يدوياً، إلى وصولها للشبكات الالتفافية CNN التي تتعلم ليس فقط التصنيف إنما أيضاً استخراج السمات الأساسية والمعقدة انطلاقاً من الصورة الخام، ونهاية بشبكات LSTM التي تستطيع اكتشاف الروابط والسمات الزمنية التي تربط المدخلات المتعاقبة، كما تم الحديث عن الشبكات الهجينة CNN-LSTM التي تكشف السمات المكانية والزمانية المتقاربة والمتباعدة ضمن الفيديو، وقد أُورِدَ ضمن هذا الفصل الحديث عن صورة التدفق البصري التي تعطي كمية الحركة واتجاهها وتستخدم لمعرفة السمات الزمانية القصيرة أو المتقاربة، ومع ذلك فإن هنالك صور أخرى أفضل من صور التدفق البصري في وصف الحركة ضمن الفيديو سيتم الحديث عنها في الفصل التالي.

الفصل الثاني

صورة تاريخ الحركة Motion History Image

2-1- تمهيد

لم يتوقف الاهتمام في مجال الإبصار الحاسوبي على تقنيات تحليل الصور، فمع تطور تلك التقنيات اتجه الباحثون نحو تحليل الحركة وكشف السلوك أو النشاط البشري [68]، فقد ظهرت في ثمانينات القرن الماضي طريقة التدفق البصري لتحليل وكشف الحركة، إلا أنها تحتاج عمليات حسابية معقدة ووقت معالجة طويلاً كما أنها حساسة للضجيج [69]، فهي تعتمد على حساب مركبتي السرعة الأفقية والשאقولية بين الإطارات المتعاقبة علاوة على حساب تغير شدة البيكسل مع الزمن [70]، كما أنها مناسبة فقط لتحليل الحركات البطيئة. في الحقيقة كان الاهتمام منصّباً على تحليل حركات اليد لبناء برامج تفاعلية بين الإنسان والحاسوب [71, 68]، مما دعا Davis وآخرون [72] إلى تطوير طريقة تعتمد على العرض أو المظهر (View-based approach) في التعرف على النشاط مباشرة من الحركة ذاتها مستمدة من طريقة مقترحة لكشف حركات الوجه [73]، وقد أطلقوا عليها صورة تاريخ الحركة Motion History Image MHI. فالطرق التي تحلل وتكشف الحركة هي على ثلاثة أنواع: الأول يسمى استعادة النموذج البشري العام (Generic human model recovery) الذي يعتمد على مقارنة مسقط شكل الشخص مع نموذج ثلاثي الأبعاد مخزن مسبقاً لنماذج الحركات وذلك من أجل كل لحظة زمنية، أما النوع الثاني فهو طرق تعتمد على المظهر، والتي تعتمد على مقارنة الصور الثابتة المتعاقبة المستخرجة من الفيديو مع قوالب محفوظة مسبقاً لتمييز النشاط البشري، أما الثالث فهو طرق تعتمد

على الحركة مباشرة (Direct motion-based approaches)، والتي تعتمد على خواص الحركة ذاتها المستنتجة من الصور المتعاقبة لكشف النشاط البشري.

2-2- صورة تاريخ الحركة MHI وصورة طاقة الحركة MEI

يُعرّف الانتقال على أنه الحركة مع مرور الزمن، ومجموعة الانتقالات تعرّف نشاطاً معيناً ما [4]، أوحث هذه الفكرة إلى كل من Davis و Bobick الاعتماد على الطريقة المرتكزة على المظهر لمطابقة القالب؛ بمعنى أنه لا تُعامل كل صورة على حدى ضمن تسلسل الصور، بل يتم تكثيف تسلسل صور الهيئة (Silhouette) وفق تابع معين بحيث يُعبّر عن ذلك التسلسل في صورة واحدة ذات تدرج رماديات تعبر عن قالب من خلاله يكشف النشاط البشري المحدد، وبهذا تمثل الصورة الأخيرة الحركة في حالة مدمجة، سميت تلك الطريقة قالب صورة تاريخ الحركة MHI.

لقد كانت تلك الطريقة رداً على تساولين: الأول أين (Where) حدثت الحركة ضمن الصورة، أما الثاني فهو كيف (How) حدثت تلك الحركة، أما جواب التساؤل الأول فكان من خلال صورة طاقة الحركة (Motion Energy Image MEI) التي تصف شكل الحركة والتوزيع المكاني لها [4, 72]، بينما تقوم MHI بتحديد كيف حدثت الحركة، حيث شدة كل بيكسل فيها تابع لكثافة الحركة في ذلك المكان من صورة MHI، وباستخدام كلا الصورتين يمكن التنبؤ تماماً بالفعل أو النشاط الذي قام به الشخص.

تتلخص صورة تاريخ الحركة MHI في أخذ صورة الفرق $D(x, y, t)$ بين هئيتي الشخص في إطارين متتابعين أو على بعد خطوات محددة ضمن الفيديو ثم إخضاعها للمعادلة التالية لتوليد صورة H_t MHI:

$$H_t(x, y, t) = \begin{cases} \tau & \text{if } D(x, y, t) = 1 \\ \max(0, H_t(x, y, t-1) - 1) & \text{otherwise} \end{cases} \quad (1-2)$$

حيث تشير τ إلى عدد بين 1 و 255 يمثل الامتداد الزمني للحركة، فتكون H_t عبارة عن صورة فيها شدة البيكسل في كل نقطة منها هي تابع زمني لمراحل الحركة في تلك النقطة، وبذلك تكون البكسلات التي

تتتمي للإطارات¹ الأحداث التي حدث فيها التغير أكثر سطوعاً من تلك للإطارات القديمة ضمن الصورة MHI، إذ تضمحل قيم بيسكلات الإطارات الأقدم مع مرور الزمن. يبين الشكل (1-2) صورة الشخص بعد أن رفع ذراعيه وصورة MHI المرافقة.



a- arms-wave



b-arms-wave MHI

الشكل 1—2 a- صور شخص يلوح بذراعيه، b- صورة MHI [4].

يمكن تعميم المعادلة (1-2) لتصبح على الشكل التالي:

$$H_{\tau}(x, y, t) = \begin{cases} \tau & \text{if } \Psi(x, y, t) = 1 \\ \max(0, H_{\tau}(x, y, t - 1) - \delta) & \text{otherwise} \end{cases} \quad (2-2)$$

حيث يشير تابع التحديث $\Psi(x, y, t)$ إلى وجود الحركة في صورة الفيديو الحالي، يمكن أن يكون هذا التابع هو طرح الصورة من الخلفية أو الفرق بين صورتين هينتين متعاقبتين أو التدفق البصري [74]، بينما تمثل δ معامل الاضمحلال، عادة ما يكون تابع التحديث هو الفرق بين صورتين متعاقبتين للشخص المراد كشف نشاطه، حيث تعطى علاقة تابع التحديث بالمعادلة التالية (2-3).

$$\Psi(x, y, t) = \begin{cases} 1 & D(x, y, t) \geq \xi \\ 0 & \text{otherwise} \end{cases} \quad (3-2)$$

¹ الإطار هو صورة مستخلصة من الفيديو في لحظة زمنية ما t.

حيث التابع $D(x,y,t)$ هو تابع الفرق بين إطارين متعاقبين بفارق زمني Δ ويعطى بالمعادلة (4-2).

$$D(x,y,t) = |I(x,y,t) - I(x,y,t \pm \Delta)| \quad (4-2)$$

تعبّر $I(x,y,t)$ عن شدة البكسل في الموضع (x,y) من الإطار ذي الترتيب الزمني t .

أما صورة طاقة الحركة MEI التي ذكرناها آنفاً، فهي عبارة عن صورة حركة ثنائية تراكمية تصف كيف تمت الحركة في تسلسل الفيديو، وتحسب من إطار البداية وحتى النهاية في ذلك الفيديو، والجسم المتحرك في الفيديو تغطي حركته مساحة أو منطقة ما ضمن إطار الفيديو، وشكل هذه المنطقة يحدد أين حدثت الحركة [75]؛ وهذا ما تعبر عنه المنطقة المضاءة في صورة طاقة الحركة، تعطى صورة MEI من خلال المعادلة (2-5).

$$E_{\tau}(x,y,t) = \bigcup_{i=0}^{\tau-1} D(x,y,t-i) \quad (5-2)$$

يمكن استنتاج صورة MEI هذه من صورة MHI من خلال تطبيق عتبة محددة على تلك الصورة

ويمكن التعبير عن ذلك من خلال المعادلة (2-6) التالية.

$$E_{\tau}(x,y,t) = \begin{cases} 1 & H_{\tau}(x,y,t) \geq 1 \\ 0 & \text{otherwise} \end{cases} \quad (6-2)$$

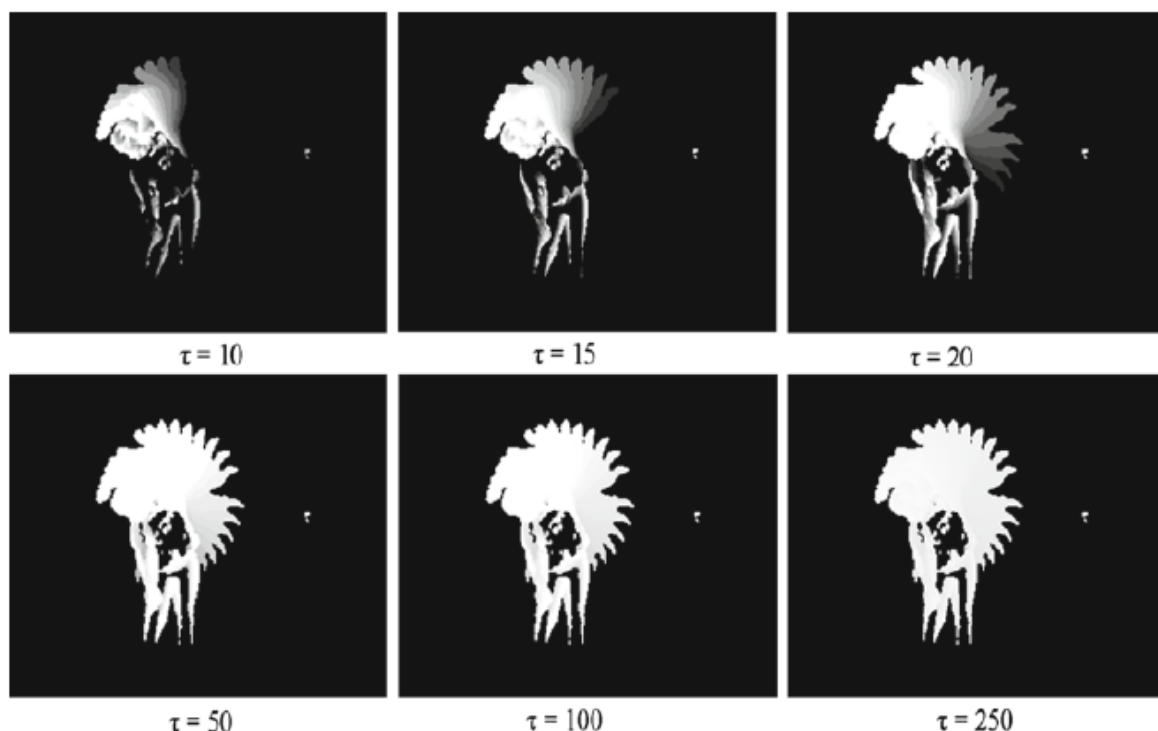
فكما هو ملاحظ في صورة MHI فإن تدرج الرماديات يفيد في تحسس اتجاه الحركة وبالتالي التمييز بين النشاطات التي تحدث باتجاهات متعاكسة (مثل "القيام" مقابل "الجلوس")، وعليه يجب أن تستخدم كل من صورة MHI و MEI حتى يتم تمثيل الحركة بكل معلوماتها، فاستخدام كلا الصورتين معاً يقدم تمييزاً أفضل للنشاط البشري من استخدام كل منهما على حدا [4]. ويبين الشكل (2-2) نشاطين متعاكسين وصورتين كل من MHI و MEI لهما، حيث أحدهما ينحني ملوحاً بيده اليسرى صعوداً والآخر يعمل حركة معاكسة ولكن بيده اليمنى، يلاحظ أنه لو أخذت صورة MEI لأحدهما وعكست وفق المحور الشاقولي لنتجت صورة MEI الأخرى، وهذا لعدم قدرة MEI تمثيل كيف تمت الحركة.



الشكل 2-2 الصف الأول الأعلى من اليسار يمثل أربعة إطارات من فيديو لرجل ينحني رافعاً يده اليسرى، أما الصورة الخامسة فهي صورة MHI الممثلة لتلك الحركة، والسادسة هي صورة MEI، أما الصف الثاني يمثل أربعة إطارات لرجل ينحني نزولاً باستخدام يده اليمنى، والصورة الخامسة هي صورة MHI لهذه الحركة، أما السادسة فهي صورة MEI [69].

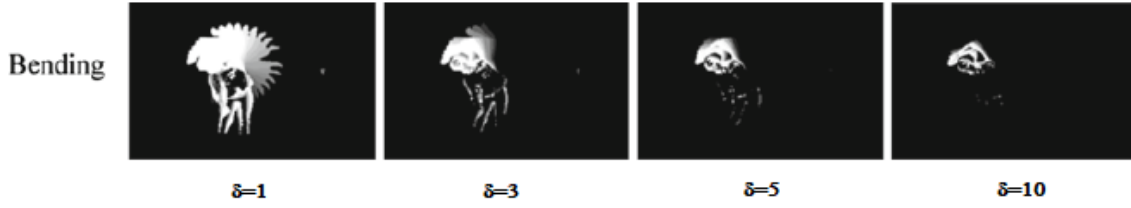
2-2-1- تأثير المعاملان τ و δ

بالعودة للمعادلة (2-2)، فإن للمعامل الزمني τ تأثيراً كبيراً على صورة MHI، فلو قل ذلك المعامل عن عدد الإطارات المكونة للنشاط المطلوب تمييزه، فإن معلومات الإطارات السابقة ستضيع عند قدوم الإطار الجديد؛ على سبيل المثال لو كان مشهد النشاط مؤلفاً من 26 إطاراً وكان $\tau=15$ ، فإنه عند قدوم الإطار رقم 15 فإن معلومات الإطار الأول ستضيع وتضمحل، وذلك عندما تكون قيمة معامل الاضمحلال $\delta=1$ ، أما عندما تكون قيمة المعامل الزمني τ كبيرة مثلاً 255 بالنسبة لذلك المشهد فإن قيم التغير بين الأطر ستكون ضئيلة جداً وستبدو صورة MHI مشابهة إلى حد كبير لصورة MEI، لذلك يجب اختيار قيمة τ بعناية عند حساب صورة MHI، يبين الشكل (3-2) صور MHI المقابلة لقيم مختلفة للمعامل τ لذات النشاط في الشكل (2-2).



الشكل 3-2 صور MHI لذات المشهد ولكن بقيم τ مختلفة [69].

أما بالنسبة لمعامل الاضمحلال δ -القيمة الافتراضية له 1 في المعادلة (2-1)- فإن تغييره أيضاً يلعب دوراً في بقاء أو اختفاء المعلومات المتعلقة بالحركة، فأتثناء عملية حساب قيمة بيكسل في صورة MHI في أحد الإطارات ولم يحدث تغير في قيمة ذلك البيكسل عن الإطار السابق؛ أي أنه لا توجد حركة، فإنه ووفقاً للمعادلة (2-2) ستتقص قيمة ذلك البيكسل بمقدار δ فإن كان الناتج أصغر من الصفر فإن القيمة الجديدة ستكون صفراً وبالتالي سيضمحل البيكسل وتضيع معه معلومات الحركة السابقة، فالقيمة الكبيرة للمعامل δ تحذف المسار السابق في تسلسل الحركة، لذلك يجب على الباحث أن يختار هذا المعامل بعناية عند حساب صورة MHI، ويبين الشكل (2-4) صور MHI المحسوبة لذات النشاط المبين في الشكل (2-2) وذلك من أجل قيم δ مختلفة ($\delta=1, 3, 5, 10$).



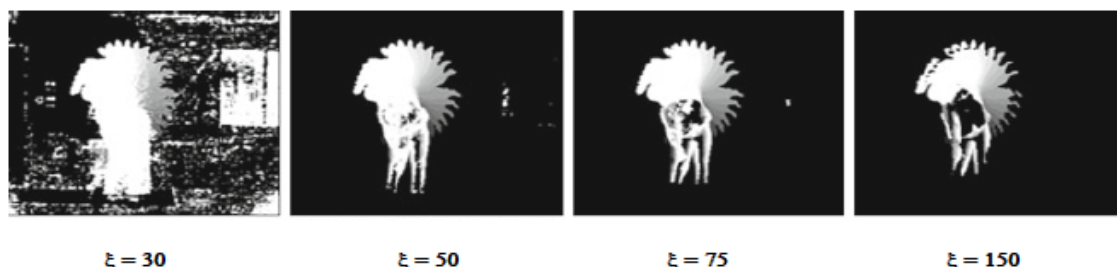
الشكل 2-4 صور MHI لنشاط الانحناء ولكن بقيم مختلفة لـ δ [69].

نرى مما سبق أن معلومات الحركة في الأطر السابقة المكونة لصورة MHI هامة، ويجب تعديل كل من قيمتي المعاملين τ و δ عند انتاج صور MHI، لأنه من الواضح تماماً أنهما معاً سيحددان مسار الحركة ومتى سيضمحل في صورة MHI، وبالتالي يحددان النافذة الزمنية التي تؤخذ خلالها صورة MHI؛ فمثلاً اختيار تركيبة هذان المعاملان حيث τ كبيرة و δ صغيرة يؤدي إلى صورة MHI فيها الحركة عبارة عن تدرج مستمر متغير ببطء، بينما التركيبة التي يكون فيها كلا المعاملين كبيراً ستؤدي إلى صورة MHI ذات حركة مكتملة متقطعة بخطوات.

2-2-2- تأثير تابع التحديث Ψ ومعامل تحسس الحركة ξ

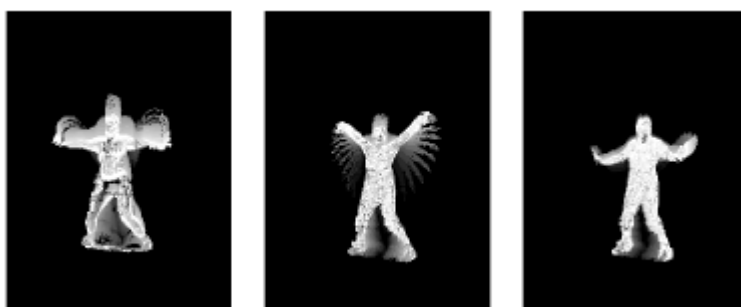
يلعب تابع تحديث الحركة $\Psi(x,y,t)$ دوراً هامة في حساب صورة MHI، فحسب ذلك التابع يختلف شكل صورة MHI الناتجة، يستخدم تابع تحديث الحركة الأساسي -المعادلة (2-1)- المعتمد على الفرق بين هئيتي الشخص في إطارين متعاقبين [76-79] بشكل واسع، ويقدم طريقة جيدة لاستخلاص صورة الحركة ومتكيف مع البيئة الديناميكية، إلا أن عدم التوفيق في اختيار قيمة معامل التحسس يمكن أن يتسبب بثقوب سوداء (فقدان الحركة) ضمن الجسم المتحرك وثقوب بيضاء (ضجيج) ضمن البيئة المحيطة، حيث يبين ذلك الشكل (2-5) إذ يلاحظ تشكل نقاط عديدة بيضاء في خلفية صورة MHI عندما اختير معامل التحسس صغيراً جداً $\xi=30$ ، حيث يبين الشكل صورة MHI المحسوبة بقيم مختلفة للمعامل ξ ، أما زيادة ذلك المعامل فسيؤدي أيضاً مع اختفاء الضجيج إلى ضياع أجزاء من معلومات الحركة، حيث نلاحظ أن الجسم متحرك أيضاً ولكن

مع زيادة قيمة ξ فإن أجزاء منه تختفي، كما هو واضح في الصورة الرابعة أقصى اليمين ولكن الحركة الأقوى وهي التلويح باليد بقيت.



الشكل 2-5 أهمية اختيار معامل التحسس أو العتبة بالنسبة لتابع التحديث [69].

إن الاختيار الصحيح لتابع التحديث يلعب دوراً كبيراً في الحفاظ على معلومات الحركة، فعندما يكون هذا التابع بسيطاً كما في [4] فإن صورة MHI ستحتوي على ضياع في معلومات الحركة لبعض أجزاء الجسم، فكما يبين الشكل (2-6) فإن الفتاة التي تقوم بحركات أيروبيك تحرك جسمها ببطء بينما تلبس لباساً ذو تردد عالي (مبرقع، ذو تفاصيل كثيرة) مما انعكس على فقدان في أجزاء من الجسم المتحرك أثناء القيام بالحركة وهذا ما يظهر على شكل مناطق سوداء ونقاط سوداء ضمن جسم تلك اللاعبة في صور MHI المستنتجة لها.



الشكل 2-6 صور MHI لفتاة الأيروبيك تحرك جسمها ببطء [4].

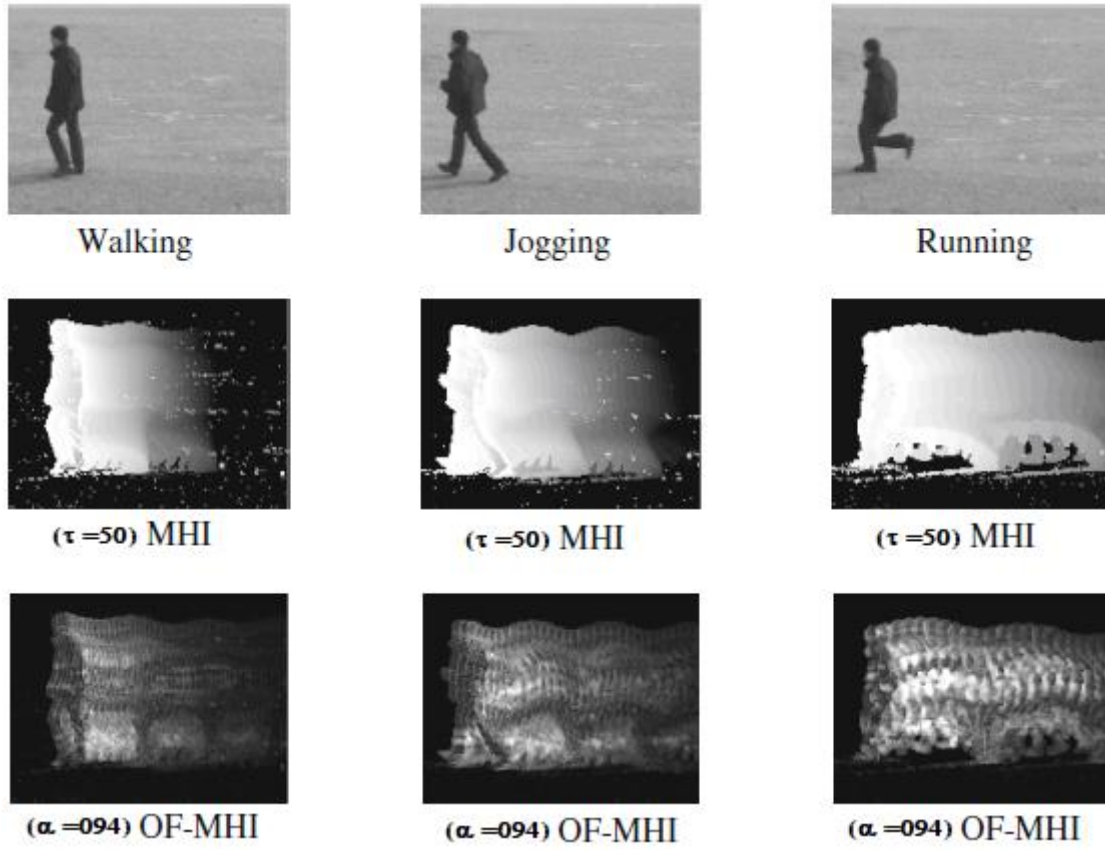
لذلك استخدمت توابع أخرى للحصول على صور MHI اعتمدت أغلبها على التدفق البصري، فقد اقترح Tasi [80] صورة MHI تعتمد على التدفق البصري سماها OF-MHI حيث وصفت معلومات الحركة بدقة عالية مقارنة بصورة MHI التقليدية، وتحسب هذه الصورة وفق المعادلة (2-7):

$$E(x, y, t) = s(x, y, t) + E(x, y, t - 1) \times \alpha \quad (7-2)$$

حيث يمثل التابع $E(x, y, t)$ قيمة البيكسل في صورة OF-MHI الجديدة، و α هو معامل معدل التحديث ويأخذ قيم بين الصفر والواحد، أما $s(x, y, t)$ فيمثل التدفق البصري في الموضع (x, y) في اللحظة t ، ويعطى بالمعادلة (8-2).

$$s(x, y) = \sqrt{u^2(x, y) + v^2(x, y)} \approx |u(x, y)| + |v(x, y)| \quad (8-2)$$

إذ تمثل $s(x, y)$ مركبة السرعة الكلية المحسوبة في الموضع (x, y) - مع التتويه بأن معادلتني حساب كل من سرعتين u الأفقية و v الشاقولية معقدتان جداً - وعليه فإن قيمة كل بيكسل في الصورة الجديدة يدخل في حسابها مركبة السرعة تلك، وبالتالي فالحركات البطيئة تكون البكسلات لها قيم قليلة مقارنة بالحركات السريعة على عكس MHI التقليدية التي تأخذ فيها البيكسل القيمة ذاتها τ مهما كانت طبيعة الحركة، أما α فإنه يراكم قيمة البيكسل السابق إلى الجديد أسياً إن كان الجديد هو جزء من هيئة الشخص (يطلق عليه الصورة الأمامية)، أو ينقص قيمة البيكسل أسياً إن كان من الخلفية، وهذا ما يظهره الشكل (7-2).



الشكل 7-2 صور كل من MHI و OF-MHI لثلاثة نشاطات مختلفة [80].

إذ يلاحظ أن صور OF-MHI تمثل الحركة بشكل أفضل من تلك في صور MHI التقليدية، وعلى الرغم من ذلك فإن صور OF-MHI تتطلب حسابات معقدة وتستهلك موارد الذاكرة، وتتأثر بالضجيج وتعقيد الخلفية، ولكن هذا لا يظهر في الشكل (7-2) كون الخلفية المختارة بسيطة، والصور نقية لا تحتوي على ضجيج.

3-2- قصور طريقة MHI

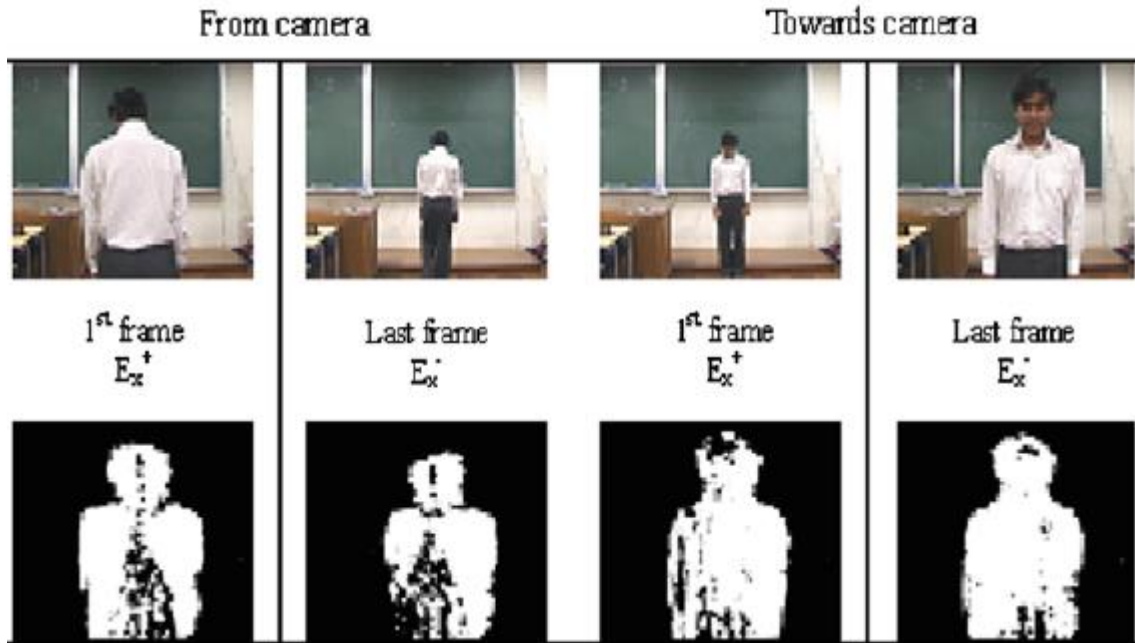
على الرغم من النجاح الذي أحرزته طريقة MHI في تمييز النشاط البشري، إلا أن لها العديد من العيوب التي يمكن إجمالها فيما يلي:

هذه الطريقة ليست مناسبة للخلفيات الديناميكية، وخاصة مع الطريقة التقليدية لاستخراجها المعتمدة على تابع الفرق بين هئيتي الشخص في إطارين متعاقبين وطرح الخلفية [81]، ولكن مع استخدام طرق تستطيع استخلاص هيئة الشخص من الخلفية الديناميكية يمكن لهذه الطريقة أن تكون فعالة.

إن طبيعة طريقة MHI لا تساعد في تتبع مسارات الأشياء إن أريد ملاحقة تلك الأشياء كالأشخاص والسيارات [82].

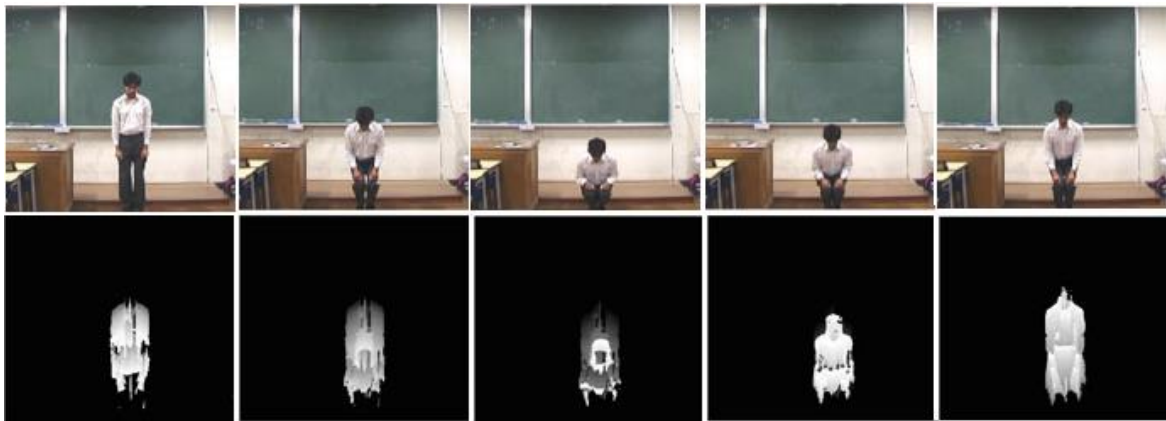
كما بالنسبة للخلفية الديناميكية يجب أن يكون في المشهد فقط شيء واحد يتحرك، أما إن وجدت أشياء أخرى قابلة للحركة فيجب أن تكون ساكنة وإلا أدى ذلك إلى فشل تمييز النشاط [83].

إن العيب الرئيسي في طريقة MHI التغطية الذاتية (self-occlusion) وهي أن الأشياء التي تتحرك تغطي أشياء قد تحركت قبلها وبالتالي فالحركة الحديثة تمحو أو تكتب فوق القديمة وبالتالي تغطيها فتختفي تلك المعلومات المتعلقة بالحركة السابقة، ويظهر هذا العيب بشكل واضح عندما يتحرك الشخص في مواجهة الكاميرا ذهاباً أو إياباً (أي أحد محاور الحركة موازي لمحور الكاميرا) كما في الشكل (2-8)، وكذلك عندما يقوم الشخص بحركة ثم يقوم بحركة معاكسة ("وقوف" ثم "قرفصاء" ثم "وقوف") كما في الشكل (2-9).



الشكل 2—8 العامودان على اليسار لشخص يبتعد عن الكاميرا مع صورتي MEI بينما على اليمين لشخص يقترب من

الكاميرا مع صورتي MEI، حيث تفشل MHI و MEI في تمييز النشاطين [69].



الشكل 2—9 شخص يقوم بالقرصاء ثم بعد ذلك يعود إلى الوقوف وصور MHI الموافقة [69].

تعاني طريقة MHI من عيب آخر هو حساسيتها للتغيرات في أزمنة الحركة، فقد لا تستطيع التمييز بين الجري والركض إن كانت السرعات متقاربة [84]، إضافة لحساسيتها لسرعة الفيديو.

أما العيب الأخير فهو من الطريقة ذاتها، إذ أنها مخصصة لكشف الحركة فقط في جسم الشخص وبالتالي الأجزاء الساكنة من الجسم لن يتم تمثيلها، فهناك نشاطات تقدم فيها الأجزاء الساكنة من الجسم معلومات عن

طبيعة النشاط الذي يقوم به الشخص، بل إن هنالك أوضاعاً مرحلية في أثناء الحركة يكون فيها أحد أطراف الجسم ساكناً لفترة ولكنه يحتوي معلومات عن طبيعة الحركة، وهذا ما أغفلته MHI، فمثلاً حدث السقوط كما صنفته Frassetto [26] يقسم إلى نوعين سقوط غير مباشر وفيه يتعثر الشخص بشيء فيتحرك جسده كتلة واحدة تقريباً إلى حين الارتطام بالأرض ثم يأخذ وضعيات مختلفة يتخللها أجزاء ساكنة من جسمه تبعاً للسقوط إلى الأمام أو الخلف أو أحد الجانبين، والنوع الآخر هو السقوط المباشر وفيه يسقط مغشياً عليه وتكون حركته عامودية ثم يتوقف بفعل ارتطام ركبته بالأرض أو إلبته حيث يتعرض لتوقف لبرهة خلالها تتغير طريقة السقوط إلى ما يشبه السقوط غير المباشر، وهذه الوضعيات المرحلية الساكنة تلعب دوراً في تحديد طبيعة الحركة، وقد أهملتها صور MHI، كما يلعب التلاشي أو الاضمحلال للإطارات القديمة دوراً سلبياً أيضاً في ضياع معالم بداية الحركة.

4-2- منهجيات استخدمت فيها MHI

استخدمت MHI بشكل مباشر بدون تعديلات في العديد من المنهجيات [87-85,83,36,34,25] التي تُستخدم لكشف النشاط البشري، وتستخدم كذلك في تمييز الإيماءات لمساعدة المعوقين [88] وفي تمييز الصور التي تحوي على كلام ضمن الفيديو حيث تبقى على حركة الفم فقط [90,89].

استخدمت MHI بعد إجراء بعض التعديلات عليها وكان صاحب النصيب الأكبر في ذلك منهجيات كشف وتمييز المشي (Gait Recognition)، حيث اعتمداً على طريقة MHI استنبط Han وآخرون [91] طريقة لتمييز المشي وأطلق عليها صورة طاقة المشي (Gait Energy Image GEI)، حيث بالمقارنة مع MHI وMEI، خصصت هذه الطريقة فقط لتمثيل حركة الشخص الماشي بشكل طبيعي والوحيد في المشهد، وتمثل المعادلة (9-2) كيفية حساب هذه الصورة:

$$G(x, y) = \frac{1}{N} \sum_{t=1}^N B_t(x, y) \quad (9-2)$$

حيث N هو عدد الإطارات المكونة لدورة مشي كاملة، بينما $B_t(x,y)$ هي قيمة البيكسل في الموضع (x,y) لهيئة الشخص الثنائية في الزمن t ، ومن هذه المعادلة يتبين أن GEI هي عبارة عن صورة طاقة تراكمية زمنية قياسية، مع أنها فعالة في تمييز المشي إلا أنها ليست بكفاءة طريقة MHI عندما يتعلق الأمر بكشف بقية النشاطات البشرية، لقد ظهرت لهذه الطريقة في تمييز المشي تعديلات وتطويرات عديدة [95-92].

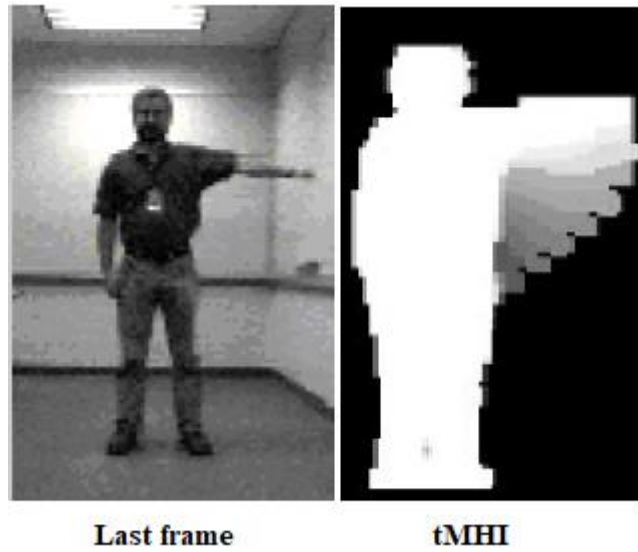
أما في السنوات الأخيرة فقد استفيد من صورة MHI في كشف السقوط - فكشف هذا النشاط حطي بدراسات عدة في العقد الأخير - وذلك باستخدام تقنيات التعلم العميق، فهناك دراسات استخدمت صور MHI كما هي مع الشبكات العصبونية CNN كما فعل Kong [36]، وكذلك Lie [34]، أو بعد إجراء تعديل عليها كما فعلت Khraief [35,11].

2-5- مناهج أخرى مطورة لطريقة MHI

بهدف تلافي عيوب طريقة MHI ظهرت العديد من الطرق التي أساسها MHI ، كان أولها مقترح من قبل مخترعي MHI وهما Bradski و Davis [96]، حيث عدلا الطريقة بحيث تتلافى سيئة حساسيتها لمعدل الإطار (ضمن حدود معينة) وقد أطلقا على تلك الطريقة ($tMHI$ Image Motion History Image) وتعطى بالمعادلة (2-10):

$$tMHI_{\delta}(x,y) = \begin{cases} \tau & \text{if current silhouette at } (x,y) \\ 0 & \text{else if } tMHI_{\delta}(x,y) < (\tau - \delta) \end{cases} \quad (10-2)$$

حيث τ هو الطابع الزمني الحالي - وليس رقم الإطار - ويعطى بالفاصلة العائمة، بينما δ هو ثابت الفترة الزمنية الأعظمية (نموذجيا عدة ثواني)، وبهذه الطريقة حركة اليد أو الإيماءة ستحتل ذات المساحة في صورة $tMHI$ مهما كان معدل الإطار أو سرعة التقاط الصور، كما تقدم هذه الطريقة منهجية تجزيء الحركة اعتمادا على تجزيء مناطق الحركة إلى طبقات، وليس كما سبق منطقة الحركة كبقعة واحدة متدرجة اللون، ويبين الشكل (2-10) ذلك.



الشكل 2—10 يظهر رجلاً يرفع يده مع بقاء جسده ساكناً، كما يظهر صورة tMHI المقابلة [96].

مع أن تلك الطريقة تمثل المناطق التي حدثت فيها الحركة، ولكن الحركة تتلاشى مع تقدم الزمن، إضافة إلى أن هذه الطريقة مع أنها على عكس السابقة تظهر المناطق الساكنة في الجسم إلا أنها لا تعطي انطباعاً عن فترة بقاء تلك المناطق ساكنة.

لحل عيب التغطية الذاتية اقترح Pantic [97] طريقة تقوم على تسجيل صورة تاريخ الحركة على فترات زمنية متعددة (صور MHI متعددة المستويات)، إذ ربط قيمة كل بيكسل من إطار ما ضمن تسلسل الأطر بسوية لذلك الإطار تحسب من رقمه ضمن التسلسل منسوباً للسوية العظمى ذات القيمة 255، ودون استخدام قاعدة الاضمحلال في MHI التقليدية، وقد أطلق على تلك الطريقة تسمية (Multi-level MHI MMHI)، واستخدمها في تمييز تعابير الوجه، وتحسب صور MMHI باستخدام المعادلة (2-11):

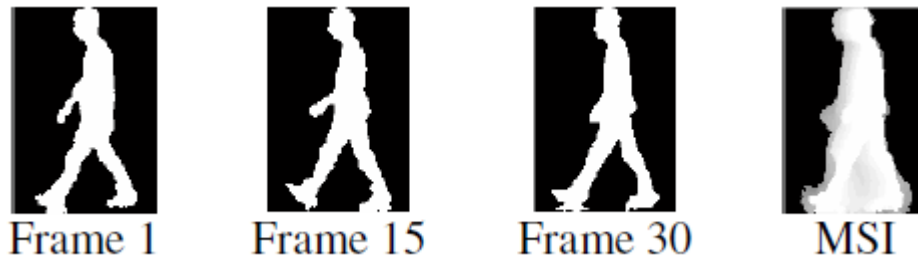
$$MMHI(x, y, t) = \begin{cases} s * t & \text{if } \Psi(x, y, t) = 1 \\ MMHI(x, y, t - 1) & \text{otherwise} \end{cases} \quad (11-2)$$

حيث $s = 255/n$ هي خطوة الشدة بين سويتي تاريخ حركة متعاقبتين، بينما $MMHI(x,y,t) = 0$ من أجل $t \leq 0$ ، بينما $t = 1..n+1$. مع أنها تلافت مشكلة التغطية الذاتية في MHI إلا أنها لم تثبت فعالية كبيرة في تمييز الحركة كما لطريقة MHI وذلك عند تجربتها على مجموعة حركات الأيروبيك [98].

قام Lam وآخرون [99] باقتراح طريقة لتلافي ضياع الأجزاء الساكنة من صورة MHI، حيث تعتمد كتابع تحديث على هيئة الشخص $B(x,y,t)$ ذاتها وليس على صورة الفرق بين هيئتين، وقد سمي الطريقة (Motion Silhouette Image MSI)، وأعطيت بالمعادلة (12-2):

$$MSI(x,y,t) = \begin{cases} 255 & \text{if } B(x,y,t) = 1 \\ \max(0, MSI(x,y,t-1) - 1) & \text{otherwise} \end{cases} \quad (12-2)$$

تشابه هذه الطريقة إلى حد كبير MHI، حيث $\tau=255$ و $\delta=1$ ، ولكن مع تضمين الأجزاء الساكنة من الجسم، وعلى الرغم من ذلك تتلاشى الأجزاء الساكنة والمتحركة مع تقدم تسلسل الإطارات، بحيث تبقى الإطارات الأحدث، ويبين الشكل (11-2) مثال على هذه الصورة لشخص يمشي.



الشكل 2—11 عدة إطارات من مشية شخص مع صورة MSI المقابلة [99].

اعتماداً على فكرة MHI في تمييز السلوك البشري، طرح Chen وآخرون [100] طريقة سماها صورة الحركة ذات الخطوة (Stride History Image SHI)، وقد اقترح صورة خطوة إلى الأمام وصورة خطوة إلى الخلف، إذ بهذه الطريقة يمكن أن يتلافى تلاشي الأطر السابقة في التسلسل عند تشكيل الصورة SHI، فالصورة الأمامية fSHI تحافظ على الحركات الأخيرة أكثر سطوعاً، بينما الصورة bSHI أو rSHI تحافظ على الحركات الأولى أكثر سطوعاً، وبالتالي يُمكن استخدام كلا الصورتين من التغلب على التلاشي، يمكن من خلال هاتين

الصورتين وصف التغيرات بين الأطر، وإعطاء المعلومات الزمنية والمكانية الخاصة بالحركة أو النشاط، تعطى

هذه الصور وفق المعادلتين (13-2 و 14-2):

$$fSHI(x, y) = \sum_{n=1}^{N_\lambda-1} fSHI(x, y, n), fSHI(x, y, n) = \begin{cases} n\tau & \text{if } D(x, y, n) = 1 \\ 0 & \text{otherwise} \end{cases} \quad (13-2)$$

$$bSHI(x, y) = \sum_{n=1}^{N_\lambda-1} bSHI(x, y, n), bSHI(x, y, n) = \begin{cases} n\tau & \text{if } D(x, y, n) = -1 \\ 0 & \text{otherwise} \end{cases} \quad (14-2)$$

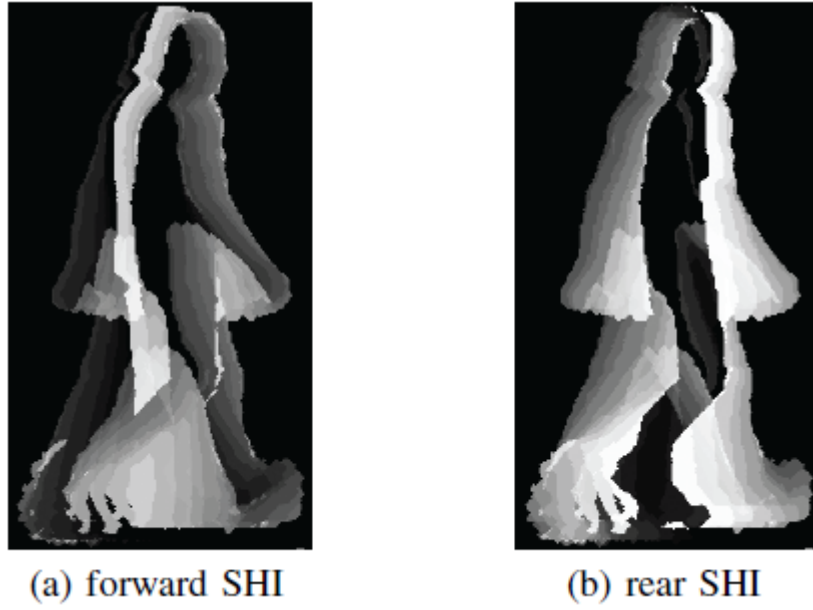
حيث $D(x, y, n)$ هو تابع الفرق بين شدتي بيكسل في إطارين متعاقبين لهما ذات الموضع (x, y) ، بينما

N_λ عدد الإطارات الكلي في فترة المشهد المراد إجراء صورة SHI له وهنا يقابل مقدار خطوة مشي $(\lambda=1,2)$ ،

بينما τ تعبر عن شدة الرمادية في المشهد وهي تساوي $255/(N_\lambda - 1)$ ، ويبين الشكل (12-2) كلاً من

هاتين الصورتين لرجل يمشي، حيث نلاحظ تراكب الصور في الحالتين بحيث تبقىان في ذات الموضع ولا

انتشار لمنطقة المشي على مساحة واسع تكافئ انسحاب الرجل أفقياً كما في صورة MHI.



الشكل 2—12 a- صورة fSHI لرجل يمشي، b- صورة bSHI لذات الرجل [100].

للتخلص من حالة التغطية الذاتية التي فيها تحدث حركات متكررة أو متعكسة كما في الشكل (2-9)، قام Foroughi وآخرون [101] باقتراح طريقة تخزن أزمنة حدوث النشاط مع التركيز على الحركات الأخيرة فيه، وقد سمي الطريقة بصورة الحركة الزمنية المتكاملة (ITMI) Integrated Time Motion Image، وتحسب من المعادلة (2-15):

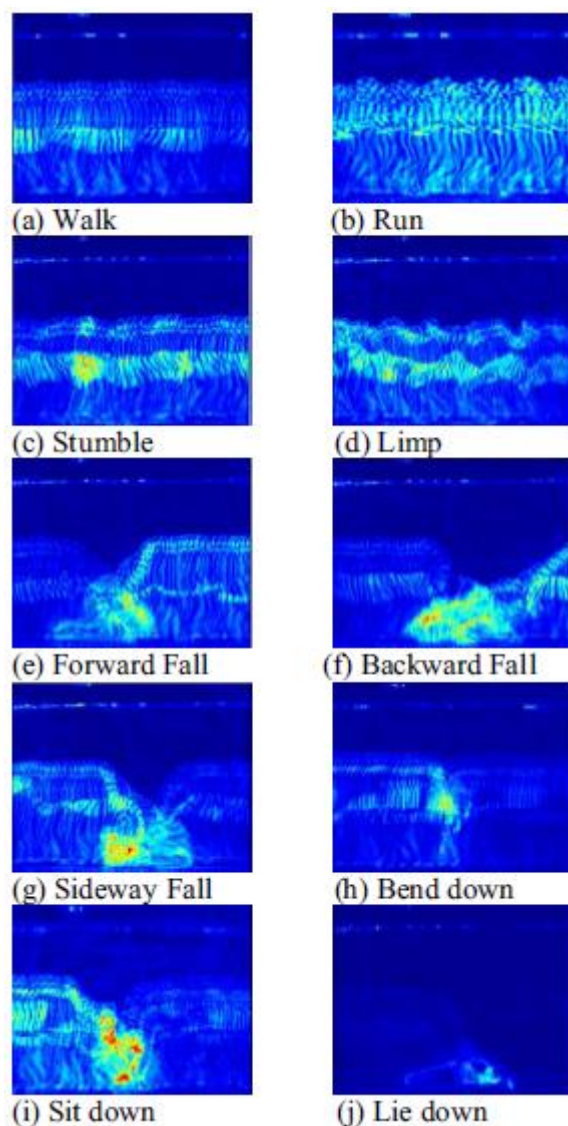
$$ITMI_t(x, y) = ITMI_{t-\Delta t}(x, y) + t \cdot f(t) \quad (15 - 2)$$

حيث $f(t)$ هو تابع الفرق ويعطى بالمعادلة (2-16):

$$f(t) = I(x, y, t) - I(x, y, t - \Delta t) \quad (16 - 2)$$

حيث $I(x, y, t)$ هي شدة البيكسل عند الإحداثيين (x, y) في اللحظة t ، وهي تشبه MHI مع تلافي

تلاشي الإطارات السابقة، ويبين الشكل (2-13) صور ITMI لمجموعة نشاطات يقوم فيها شخص.



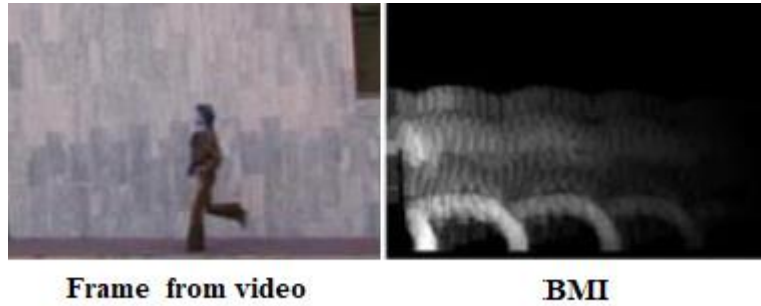
الشكل 2-13 صور ITMI لمجموعة من النشاطات التي يقوم بها شخص [101].

كما قام Doubhal وآخرون [102] بتطوير طريقة لدمج كل الصور في تسلسل النشاط في صورة واحدة أيضاً، مع التركيز كذلك على الحركات الأخيرة في النشاط، حيث استطاع تلافى تلاشي الإطارات السابقة. وقد سمي الطريقة صورة الحركة الثنائية (Binary Motion Image BMI)، وتحسب من المعادلة (2-17):

$$BMI(x, y) = \sum_{t=1}^n f(t) \cdot I(x, y, t) \quad (17 - 2)$$

حيث n عدد الإطارات المشكلة لصورة BMI، تكون فيها البكسلات الأحدث ذات إضاءة أعلى، أما $f(t)$ فهو تابع الوزن الذي يعطي إظهاراً وسطوعاً أعلى للإطارات الأحدث، وقد اختير مساوياً t^2 ليعطي نتائج

أفضل، إذ يقدم تابع الوزن طريقة لإيضاح تدفق الحركة التي تحدث ضمن النشاط أو بعبارة أخرى التدفق البصري، وعليه تستطيع هذه الطريقة تمثيل الأبعاد الزمنية والمكانية. ويبين الشكل (2-14) صورة BMI لفتاة تجري.



الشكل 2-14 صورة لفتاة تجري مع صورة BMI المقابلة [102].

كما قامت Khraief وآخرون [35] بتقديم طريقة لتشكيل صورة تدمج جميع الصور في تسلسل النشاط المراد كشفه، إلا أن المعادلة التي طرحتها هي ذات طريقة Doubhal، مع اختلاف أنها لم تحدد تابع الوزن $f(t)$ ، وقد سمت الطريقة (History of Binary Motion Image HBMI) وتحسب وفق المعادلة (2-18):

$$HBMI(X, Y) = \sum_{t=1}^n f(t) \cdot M_{xy}(t) \quad (18 - 2)$$

حيث $M(t)$ تمثل الصورة الثنائية لهيئة الشخص دون الخلفية.

2-6- الخلاصة

لقد تم الحديث في هذا الفصل عن طريقة MHI في تمثيل الحركة، حيث تم مناقشة العوامل أو المتوسطات التي تؤثر في هذه الطريقة، كما ذكرت عيوب هذه الطريقة، وأين استخدمت تلك الطريقة، ثم اختتم الفصل بالحديث عن المنهجيات المطورة اعتماداً على هذه الطريقة، حيث لوحظ بأن الطرق السابقة جميعها متشابهة في إعطاء البكسلات الأحدث في الحدث القيم الأعلى مع عدم الاهتمام بالمواضع الساكنة أو الساكنة مرحلياً التي يمر بها الشخص أثناء الحركة، وإن كان الاهتمام بها محسوساً في بعض الطرق إلا أنه لا يراعي الزمن الذي بقي فيه هذا الجزء من الجسد ساكناً.

الفصل الثالث

الدراسة المرجعية

3-1- تمهيد

يعتبر حدث السقوط من الحركات غير الطبيعية التي إن تعرض لها الإنسان قد تسبب له أذى كبيراً، قد يكتفي بكسر في أحد الأطراف أو يتعدى ذلك متسبباً بالموت، وهو يشمل جميع الأعمار ولكنه أكثر خطورة لدى الصغار وكبار السن؛ لأجل ذلك ظهرت العديد من الدراسات في العقد الأخير لتطوير أنظمة ذكية يمكنها اكتشاف حدوث السقوط [103-105]، وفي هذا الفصل سيتم الحديث عن تلك الدراسات وتقسيماتها.

3-2- أنواع طرق كشف السقوط

تصنف طرق كشف السقوط تبعاً لنوع الحساسات المستخدمة في عملية كشف السقوط إلى:

3-2-1- طرق تعتمد الحساسات المحيطية

يعتمد هذا النوع من الطرق على حساسات مثبتة في المنزل أو الأماكن التي يمشي أو يوجد فيها المرضى، إذ بالاعتماد على حساسات ضغط هوائية بيزو كهربائية قام Chaccour وآخرون [12] بصناعة سجادة مليئة بتلك الحساسات، لمراقبة كبار السن ومعرفة سقوطهم اعتماداً على قياس عتبة تجريبية للضغط وقد أعطت الطريقة حساسية مقدارها 88.8% ونوعية 94.9%.

كما قام Alwan وآخرون [13] باستخدام حساسات بيزو كهربائية ولكن لقياس نموذج الاهتزاز المتولد نتيجة السقوط وقد أعطت الدراسة حساسية وصلت إلى 100%.

قام Liu وآخرون [106] باستخدام حساسي رادار دوبلر أحدهما في السقف والآخر متوضع على

الأرض، ثم قام بحساب تحويل فورية في الزمن القصير Short-Time Fourier Transformation STFT

وحسب منه طاقة التردد مع الزمن أو ما يسمى بالمخطط الطيفي Spectrogram ومنهما استخراج مجموعة سمات بواسطة طريقة (Mel-frequency cepstral coefficients (MFCC)، بعدها استخدم تلك السمات في تعليم مصنف Support Vector Machine SVM لكشف السقوط.

على الرغم من أن هذه الطرق فعالة وجيدة إلا أنها مكلفة وتحتاج بيئة خاصة من أرضية وأسطح معينة ترتبط بها الحساسات، إضافة إلى أنها لا تصلح إلا للأماكن الداخلية (الغرف)، وغالباً ما تولد إنذارات خاطئة عندما يتم إزاحة أو تحريك أغراض ضخمة كالسرير أو الأريكة، وقد تتعرض للتشويش من البيئة المحيطة كأعمال الترميم والإنشاء في الشقق المجاورة.

3-2-2- طرق تعتمد الأجهزة الملاصقة للجسم

هي الطرق التي تعتمد على أجهزة من الممكن أن ترتبط بأماكن محددة من جسم الإنسان وتحتوي حساسات حركة وتسارع، وتكون مناسبة في حال كون الشخص في أماكن داخلية أو أماكن خارجية (باحات أو حدائق أو شوارع أو أسواق)، لقد أجريت دراسات عديدة في هذا المجال.

قام Cahoolessur وآخرون [8] بتصنيع نموذج أولي لجهاز يتوضع في حزام يلبس على خصر الإنسان، ويحتوي حساس تسارع ثلاثي المحاور مع متحكم صغري ووحدة GSM وGPS وزمور ويعمل المتحكم على نظام Raspberry Pi، وذلك لإعطاء معلومات عن التسارع في الاتجاهات المتعامدة الثلاثة (a_x, a_y, a_z) ، حيث استنتج من التسارعات الثلاثة أربع سمات، ثم استخدمها بواسطة خوارزمية XGBoost لكشف السقوط، وقد أعطت طريقته دقة 96% وحساسية 97% ونوعية 93%.

قام Islam وآخرون [14] بالاستفادة من حساس التسارع الموجود في جهاز الهاتف الذكي، وبنى ضمنه برنامج يأخذ قياسات التسارع على المحاور الثلاثة، ثم يرسلها إلى الحاسوب ليقوم من خلال قياس عتبات معينة -اعتماداً على قيم التسارع - بكشف السقوط، أعطت طريقته دقة 97.33%.

قام Dhole وآخرون [15] بالاستفادة من قبعة العمال وزرع فيها مجسين لاستحصال إشارات Electroencephalography (EEG)، حيث استخلص من هاتين الإشارتين ثمانى سمات، ثم استخدم مصنف الغابة العشوائي Random Forest Classifier لكشف السقوط، أعطت التجارب الحقيقية المجرة في المنزل وفق طريقته دقة مقدارها 98%.

قام Jácome وآخرون [16] باستخدام حساسات تسارع ثلاثية المحاور، واستخلص منها إشارات ذات سلاسل زمنية، ثم استخدم شبكة هجينة 1D-CNN-LSTM لكشف حدث السقوط، أعطت طريقته دقة 98.75% وحساسية 97.6% ونوعية 97.44%.

قام Thakur وآخرون [17] باستخدام حساس تسارع ثلاثي المحاور، وحسب منه زاوية الاتجاه orientation angle، ثم استخدم الصفات الأربعة كشعاع سمات ودرب به مصنف KNN مدرب بخوارزمية AdaBoost على كشف السقوط، واستخدم مجموعتي تدريب، أعطت دراسته دقة مقدارها 99.87% لمجموعة التدريب الأولى، و99.66% لمجموعة التدريب الثانية.

قام Tahir وآخرون [18] باستخدام حساس تسارع ثلاثي المحاور واستخلص منه التسارعات الثلاثة وأدخل كل تسارع على قناة مستقلة في شبكة عصبونية عشوائية هجينة (Hybrid Deep Random Neural Network HDRNN)، لتقوم بكشف حدث السقوط، وقد أعطت دقة مقدارها 92.23%.

قام Bharti وآخرون [3] باستخدام حزام يحتوي حساس تسارع ثلاثي المحاور وحساس حرارة وضغط جوي ورطوبة وحساس جيروسكوب، ثم استخلص من تلك الحساسات ست سمات، وقام بمرحلة التعليم من خلال خوارزمية الحقول العشوائية الشرطية (Condition Random Fields CRF)، وقد أعطت طريقته دقة مقدارها 95%.

قد أعطت هذه الطرق حلولاً جيدة لكشف السقوط إلا أنه يتوجب ارتداء تلك الأجهزة طوال الوقت، وهذا ما قد لا يريح الإنسان بشكل عام أو كبير السن بشكل خاص، أو أنه قد لا يتقبل ارتداء الأجهزة، علاوة على

أنه ربما ينسى ارتداءها، وقد يسبب الارتداء الطويل لتلك الأجهزة مشاكل جسدية خاصة إذا كانت على تماس مباشر مع جلد الإنسان.

3-2-3- طرق تعتمد الإبصار الحاسوبي

ظهرت هذه الطرق في السنوات الأخيرة، وذلك للعديد من الأسباب؛ فهي لا تحتاج لجهاز يلبس أو يركب على جسم الإنسان، كما يمكن استخدامها في جميع الأماكن الداخلية والخارجية، ولا تتطلب تجهيزات وحساسات خاصة، بل يستخدم فيها كحساسات جميع أنواع الكاميرات (كاميرات ملونة، وكاميرات العمق، وكاميرات حرارية... إلخ)، بسبب انخفاض تكلفة تلك الكاميرات وانتشارها الواسع أيضاً بين الناس من خلال الأجهزة الخليوية، وبسبب التطور الكبير في مجال تقنيات الحاسوب. تقسم طرق الإبصار الحاسوبي من حيث استخلاص السمات الخاصة بالسقوط من الصور إلى: طرق هندسية وإحصائية، وطرق تعتمد على تقنيات التعلم العميق التي زادت في العقد الأخير.

3-2-3-1 طرق هندسية وإحصائية

هذه الفئة تعتمد العديد من التقنيات لاستخلاص السمات من الصورة، كتحليل المشي، وتحليل الحركة، وتحليل الشكل، وهي سمات هندسية أو إحصائية يدخل العنصر البشري في تحديدها ثم تُستخدم طرق معينة لاستخلاصها، أما لتمييز حالة السقوط عن النشاط اليومي، فقد تستخدم مقارنة العتبات أو تعليم الآلة (Machine Learning) بفروعها.

اقترح Sehairi ورفاقه [20] طريقة تجمع بين تحليل خصائص الشكل (فرق المساحة، نسبة الارتفاع / العرض، اتجاه هيئة الشخص، قطر فيرييت، مسافة Procrustes) وخصائص الحركة (سرعة الرأس الشاقولية) بغية زيادة دقة كشف السقوط، ثم استخدم بعدها تقنيات تعليم الآلة بغية تصنيف السقوط عن النشاطات الحياتية اليومية للتأكد أن السمات التي استخلصها قوية وجيدة مهما كانت طرق تعليم الآلة المستخدمة في التصنيف،

أعطت الطريقة دقة جيدة جداً -99.27% باستخدام RBF-SVM و98.91% باستخدام KNN و99.61% باستخدام BPNN - إلا أنها عانت من سلبيات عدة أدت إلى تصنيف خاطئ، ومن هذه السلبيات مثلاً مشي الشخص خلف الأثاث، أو سقوطه بمحاذاة الكاميرا، كما أن سمات السرعة مرتبطة بفيزيولوجيا الشخص المستخدم في التدريب من وزن الرأس والجسم وغيرها، حيث تختلف من شخص لآخر وليست ذاتها لأجل كل الأشخاص، و هذا ما يمكن اعتباره انطباقاً زائداً، فقد اعتمد فقط على مجموعة FDD، وعلى لقطات لأحد المتطوعين فقط. أما Ko وزملاؤه [21] فقد طور طريقة يستخلص من خلالها العمق من كاميرا عادية وذلك باستخدام حقل ماركوف العشوائي، حيث يستخدم العمق لاستخراج هيئة الشخص دون اللجوء إلى طرق طرح الخلفية التقليدية، بعدها تستخلص سمات ثنائية البعد وثلاثية البعد لتمييز حالة السقوط، ثم اعتمد بعدها على عتبات تجريبية لتمييز السقوط عن النشاط الحياتي اليومي، وقد أعطت الدراسة دقة 99.1% وحساسية 97.3% ونوعية 99.5%، غير أنها تحمل بعض السلبيات أيضاً، أهمها اعتبار تدرج الشخص بعد السقوط وكأنه حالة تعافٍ من السقوط، اعتماده على مجموعة صور لشخص واحد فقط.

قام Albawendi وآخرون [19] باقتراح خوارزمية تعتمد على صورة MHI والشكل الإهليجي المحيط بالشخص وهيستوغرام الإسقاط Projection Histogram لتحديد حدث السقوط، حيث حسب من صورة MHI معامل الحركة C_{motion} ومن القطع الناقص حسب مجموعة سمات إحصائية تعتمد على العزوم، ومن هيستوغرام الإسقاط حسب التغيرات في الارتفاع والعرض، ثم قارن تلك السمات المستخلصة مع عتبات مستنتجة تجريبياً، ليحدد حدث السقوط، وقد أعطت الطريقة دقة 99.35% وحساسية 95.23% ونوعية 100%، حملت هذه الطريقة سلبيات عديدة، فاستخدام العتبات تجريبي متعلق بمجموعة التدريب التي صورها فقط، وبالتالي يمكن أن تكون العتبة مختلفة مع مجموعات تدريب أخرى، ورغم أنه استخدم عدة تقنيات لاستخلاص السمات إلا أنه لم يستطع تمثيل الوضعيات العابرة المرحلية الساكنة ونقاط الارتكاز الساكنة أثناء الحركة، فالدقة العالية تعود كون مجموعة تدريبه محدودة مما سبب انطباقاً زائداً.

قام Thummala وآخرون [22] باقتراح خوارزمية، تقوم على مقارنة عامل الحركة المستنتج من صورة MHI مع عتبة مستنتجة تجريبياً، ثم مقارنة تغير نسبة طول وعرض المستطيل المحيط بهيئة الشخص بين الإطارات المتعاقبة مع عتبة مقاسة تجريبياً، فإن تحققت العتبات اعتبر النشاط سقوطاً، أعطت هذه الخوارزمية دقة مقدارها 95.16%.

قام Worrakulpanit وزميله [23] باقتراح طريقة تعتمد على صور MHI حيث استنتج منها الانحراف المعياري لعامل الحركة، والانحراف المعياري لزاوية ميل الشكل الإهليجي المحيط بهيئة الشخص، ثم استخدم عتبتين لكل من هاتين الميزتين ليحدد حدث السقوط.

قام Lakshmi وزميله [24] باستنتاج سمات من صور MHI واستخدم مصنف SVM لكشف السقوط حيث دربه على مجموعة التدريب UMA، وقارن تلك النتائج مع سمات مستنتجة بواسطة Histogram of Oriented Gradients HOG، فوجد بأن صور MHI أفضل في كشف السقوط.

قامت Rougier وآخرون [25] باقتراح طريقة تعتمد على MHI وعلى تغير الشكل الإهليجي، حيث استخدمت صورة MHI فقط لمعرفة وجود حركة غير طبيعية وسريعة من خلال معامل الحركة C_{motion} ، بعدها استخدمت سمات تغير الشكل الإهليجي الإحصائية (تشتت زاوية الشكل الإهليجي مع الأفق، وتشتت نسبة القطر الثانوي للرئيسي) لمعرفة حدث السقوط، وذلك من خلال مقارنة تلك السمات مع عتبات تجريبية، واستنتجت بأن صور MHI أفضل لكشف الحركة من التدفق البصري، لأنه لا يهم اتجاه وسرعة الحركة بقدر ما يهمنا معرفة مقدار الجزء المتحرك من هيئة الشخص. كما أن التدفق البصري ليس مناسباً للعمل في الزمن الحقيقي لأنه يتطلب حسابات معقدة، ولتعرف إن كان الشخص متأدياً راقبت تغير الشكل الإهليجي خلال مدة 5 ثواني بعد السقوط فإن كانت تغيرات السمات أقل من المطلوب اعتبر الشخص متأدياً. أعطت الطريقة حساسية 88% ونوعية 87.5%.

قام Foroughi وآخرون [101] باقتراح طريقة للتعرف على السقوط، تتألف من صورة حركة مبتكرة مشتقة من MHI - المعادلة (2-15) - بعد ذلك عالج الصورة المستخلصة بتقليص السمات التي فيها من خلال

حساب القيم الذاتية في الفراغ eigenspace، ثم مرر الصور الناتجة بعد تحويلها إلى شعاع على شبكة عصبونية كاملة الاتصال FC متعددة الطبقات لتصنيف الحدث كسقوط أم لا، أعطت الطريقة دقة 89.99%.

أما Frassetto وآخرون [26] فقد قام بالمقارنة بين طريقتين تعمل أحدهما على صورة MHI لمعرفة حدث السقوط باستخدام ميزة واحدة وهي معامل الحركة C_{motion} فقط، والآخر يعمل على سمات مستخلصة من الشكل الإهليجي المحيط بالشخص، وأظهرت نتيجة المقارنة قوة MHI بالمقارنة مع طريقة الشكل الإهليجي، حيث أعطت دقة 74.98% للسقوط المباشر و75.69% للسقوط غير المباشر، بينما أعطت طريقة الشكل الإهليجي دقة 65.85% للسقوط المباشر و61.81% للسقوط غير المباشر، ولكن طريقة MHI أعطت أخطاء عندما كان الرجل يركض بسرعة أفقية كبيرة ثم يتوقف فجأة؛ ويعود السبب في هذا إلى استخدام مقياس واحد للدلالة على السقوط، بينما بالنسبة للطريقة الإهليجية فإن التغير في اتجاه الشكل الإهليجي سيؤدي إلى جعل النظام يكتشف حالة سقوط خاطئ.

قام Dubey وآخرون [27] باقتراح طريقة تعتمد على ثلاثة صور MHI، الأولى مستخلصة من صور ملونة RGB والثانية مستخلصة من صورة العمق بالزمن الأمامي fDMHI، والثالثة من صورة العمق بالزمن الخلفي bDMHI؛ أي من المستقبل إلى اللحظة الحالية إلى اللحظة السابقة، واستخلص من كل صورة من الثلاثة شعاعاً سباعي الخصائص Humoments، ثم دمج الأشعة الثلاثة وشكل منها شعاعاً واحداً مؤلفاً من 21 عنصراً، ثم استخدم بعدها مصنف SVM لتحديد حدث السقوط من عدم السقوط. أعطت الطريقة دقة وصلت إلى 97%.

قامت Alhimale وآخرون [7] باستخدام هيئة الشخص وإحاطتها بإطار مستطيل واستنتجت من الإطار مجموعة سمات، ثم استخدمت شبكة عصبونية كاملة الاتصال ودربتها على كشف السقوط، أعطت الطريقة دقة 94.3% وحساسية 93.8% ونوعية 94.8%.

تتلخص عيوب مجمل ما سبق من دراسات، في كونها تستخدم العتبات التجريبية، وسمات موضوعة بشرياً لا تكون كافية لتمثيل النشاط بشكل تام، كما أن الدقة العالية لنتائج بعض الدراسات تعود لكونها تستخدم

مجموعة تدريب محدودة خاصة بالبحث، مما يؤدي إلى الموائمة الزائدة، علاوة على أن تلك الدراسات لم تهتم بكشف نقاط ومحاور الارتكاز للشخص أثناء السقوط.

3-2-3-2 طرق تعتمد تقنيات التعلم العميق

كان Yu وزملاؤه [28] أول من اقترح طريقة لكشف وضعية سقوط الإنسان بواسطة CNN، اعتمدت الطريقة على شبكة عصبونية التلافية مؤلفة من طبقتين كل طبقة تحتوي ست خرائط خصائص بأبعاد 3×3 ومن ثم طبقتين عصبونيتين كاملتي الاتصال (FC) أبعادهما 128 و 64، يليها طبقة الخرج المؤلفة من أربعة مخارج للأصناف الأربعة (انحناء، واستلقاء، وجلس، ووقوف)، كان دخل الشبكة عبارة عن صورة لهيئة الشخص مستخلصة بطريقة طرح الخلفية المعتمدة على طريقة نموذج (codebook)، بأبعاد 30×30 ثنائية اللون أسود وأبيض، تقوم الشبكة المعلمة بالتمييز بين أربعة أصناف أحدها الاستلقاء، بعدها تُستخدم طريقة منطقة الأرضية للتأكد بأن الاستلقاء هو على الأرض؛ مما يعني أن الشخص قد سقط، تعاني هذه الطريقة من إعطاء أخطاء تصنيف خاصة عند استلقاء الشخص على الأرض بشكل طبيعي دون أن يكون سقط، إضافة إلى تصنيف الانحناء على أنه سقوط في عدة مواضع، كما أنها استخدمت شبكة CNN فقط لكشف الاستلقاء، بينما اعتمدت على طريقة منطقة الأرضية للحكم على الاستلقاء بأنه سقوط وبالتالي عاد للطرق الإحصائية لتحديد السقوط، كما أن مجموعة التدريب المستخدمة المصورة خاصةً ببحثه ولم يستخدم مجموعات تدريب أخرى، ولم تهتم الطريقة بتحليل الحركة بتاتاً ولا حتى بالوضعيات الساكنة المرحلية أو محاور الارتكاز والدوران أثناء السقوط، وقد أعطت طريقته دقة 99.01%.

كما اقترح Adhikari وآخرون [29] شبكة عصبونية التلافية CNN، مؤلفة من أربعة طبقات التلافية. أول طبقة تحوي 16 خريطة خصائص، والثانية 32 خريطة خصائص، أما الثالثة فتحتوي 64 خريطة خصائص، والرابعة تحوي 128 خريطة خصائص، حيث تتصل هذه الأخيرة بطبقة كاملة الاتصال FC شعاع الخصائص فيها مؤلف من 4096 خاصة، تليها طبقة الخرج المؤلفة من ستة مخارج يختص كل منها بوضعية من الوضعيات

الخمسـة المراد تصنيفها (الوقوف، والجلوس، والاستلقاء، الانحناء، والزحف)، أما المخرج السادس فقد خصصه للحالات التي ليست مما سبق وقد أطلق عليه اسم "الأخر"، وتكرر قيم هذه المخارج على التابع softmax الذي يعطي لكل صنف احتمال بين الصفر والواحد، وبنية هذه الشبكة التي اقترحها مستوحاة من شبكة VGGnet، أما صورة الدخل التي تمثل هيئة الشخص المراد معرفة وضعيته، فقد اعتمد في استخلاصها على طريقة لطرح الخلفية تجمع بين الصورة الملونة (RGB) وصورة العمق (Depth) المأخوذة من كاميرا Kinect لاستخراج هيئة الشخص. رغم أن هذه الطريقة لا تتأثر بمكان الشخص أو وضعية الكاميرا، إلا أنها اعتمدت في التدريب على مجموعة فيديو صغيرة مسجلة لأشخاص متبرعين مؤلفة من خمسة أشخاص من الجنسين فقط، كما أنها استخدمت شبكة CNN فقط لكشف الاستلقاء، بينما اقترح استخلاص سمات (مثل السرعة ونسبة الهيئة وزاوية الميل) بطرق رياضية، وذلك للحكم على الاستلقاء بأنه سقوط، أي أن آلية الحكم على السقوط عادت للطرق التقليدية في التصنيف، أيضاً لم تهتم الطريقة بتحليل الحركة بتاتاً ولا حتى بالوضعيات الساكنة المرحلية أو محاور الارتكاز والدوران أثناء السقوط، أعطت طريقته دقة مقدارها 74%، وحساسية 99%.

قام Núñez-Marcos وآخرون [31] باقتراح طريقة تعتمد على شبكة VGG-16 الالتفافية، حيث استخدم كدخل للشبكة صورة التدفق البصري، فهي تميز حركة الأجسام ضمن المشهد، كما تعمل في ذات الوقت كطريقة لطرح الخلفية، وقد تم تدريب الشبكة على ثلاث مراحل؛ مرحلة تدريب للشبكة على مجموعة الصور ImageNet حتى تكتسب الشبكة تمييزية كبيرة للعديد من الأجسام في الطبيعية حيث تحوي هذه المجموعة على ألف صنف، بعد أن تُدرَّب الشبكة على تلك المجموعة، تدخل المرحلة الثانية حيث تُدرَّب الشبكة على المجموعة UCF101 التي تحوي على 101 حركة نشاط يومي، ثم تليها المرحلة الأخيرة حيث تُدرَّب الشبكة على مجموعات المعطيات الخاصة بالسقوط (URFD و Multicam و FDD)، وقد أعطت هذه الطريقة دقة للمجموعات الثلاث 95% و 96% و 97% على التوالي، كما أعطت بالنسبة للحساسية أيضاً القيم 100% و 99% و 99% على التوالي، وبالنسبة للنوعية (Specificity) أعطت القيم التالية 92% و 96% و 97% على التوالي، سلبية هذه

الطريقة التعقيد الحسابي في حساب صور التدفق البصري المناسبة للحركات البطيئة، أما حسنتها فهي استخدام العديد من مجموعات التدريب.

قام Sangwan وآخرون [32] باقتراح طريقة لكشف السقوط تعتمد على صور كاميرا 360 درجة - تلتقط صورة بزاوية 360 درجة- وقد ركبها في منتصف سقف الغرفة موجهة نحو الأسفل بحيث تلتقط صورة كاملة لكل الغرفة، وعمل مجموعة صور تدريب منها، ثم اقترح طريقتين؛ الأولى شبكة 3D CNN، أما الثانية فهي هجينة من نوع inceptionV3-LSTM، حيث شبكة inception-V3 تمرر لها تلك الصور، والشعاع الناتج عنها الذي يحوي سمات حيزية مكانية يمرر إلى شبكة LSTM لتحديد التصنيف النهائي بين سقوط من عدمه، أعطت الشبكة الأولى دقة 95.3%، بينما الثانية كانت دقتها هي 91.62%، سلبية هذه الطريقة عدم الاهتمام بالوضعيات المرحلية الجزئية المكونة للنشاط، إضافة إلى أن الشعاع الناتج من inception-V3 كدخل لشبكة LSTM هو شعاع عمومي ولا يحمل سمات تفصيلية لوصف الخصائص الحيزية.

قامت Anishchenko [30] باستخدام شبكة Alexnet وتعليمها بطريقة نقل التعلم على حدث السقوط، حيث دربتها على صور من مجموعة التدريب FDD دون اقتطاع الخلفية، وقد أعطت دقة 99% وحساسية 93% ونوعية 99%، سلبات هذه الطريقة عدم الاهتمام بالسمات الحركية والاكتفاء بالسمات المكانية، والاكتفاء بمجموعة صور تدريب واحدة، مما يؤدي إلى الموائمة الزائدة.

اقترح Min وآخرون [107] طريقة تتألف من ثلاثة مراحل لاكتشاف سقوط الإنسان على أثاث المنزل كالسرير أو الأريكة أو الكرسي، وذلك لتمييزه عن الجلوس أو الاستلقاء الطبيعي على تلك الأشياء. المرحلة الأولى اعتمدت على شبكة Faster R-CNN التي تحلل المشهد وتقدم معلومات عن الموقع والأشياء في المشهد. تقوم المرحلة الثانية، التي هي وحدة استخراج سمات النشاط بحساب سمات النشاط للأشخاص الذين تم اكتشافهم، مثل نسبة العرض إلى الارتفاع للشكل البشري، والنقطة الوسطى، وسرعة الحركة أثناء اكتشاف الأشخاص وتعقبهم. أما المرحلة الثالثة فهي وحدة التعرف على السقوط والنشاط اليومي (ADL)، حيث تستخدم هذه الوحدة محرك قرار تلقائي وسلسلة من المعايير لتمييز السقوط عن النشاطات اليومية من خلال قياس التغيرات في هذه

الخصائص والحكم على العلاقة بين الأشخاص والأثاث. أعطت هذه الطريقة دقة مقدارها 93.5% لمجموعة التدريب التي جمعها، ودقة 94.1% لمجموعة التدريب المعيارية، لم تهتم الدراسة بالوضعيات المرحلية الساكنة العابرة أو محاور الارتكاز أثناء السقوط.

قامت Sherin P وآخرون [108] بتقديم طريقة لتمييز وضعية الأشخاص، تتألف من مرحلتين المرحلة الأولى هي الكشف عن وجود شخص وبالتحديد إنسان في الصورة، وذلك من خلال استخدام شبكة عصبونية الالتفافية تدعى MobileNet تقوم بتحديد فيما إذا كان الشيء الموجود في الصورة إنساناً أم لا، فإن كان إنساناً، تدخل الصورة في المرحلة التالية في التصنيف وهي شبكة inception v3 الالتفافية، حيث قامت بتدريب الشبكة على مجموعتين من بيانات سقوط الإنسان هما FDD و URFD، وتم التدريب بطريقة نقل التعلم، ميّزت الشبكة المدربة بين أحد أربعة أصناف هي: الوقوف والجلوس والانحناء والسقوط، وقد أعطت لمجموعة FDD دقة مقدارها 98.5% وحساسية 93.47% و نوعية 97.2%، أما بالنسبة للمجموعة URFD فقد أعطت دقة مقدارها 91.5% وحساسية 100% ونوعية 94%، رغم أنها استخدمت مجموعتي تدريب إلا أنها أغفلت السمات الخاصة بالحركة واكتفت بالسمات المكانية الحيزية فقط.

قام Salimi وآخرون [109] باستخدام شبكة 1D-CNN اقترحها، دخلها عبارة عن شعاع سمات مستخلص من طريقة fast pose method لمفاصل الجسم مأخوذة من مجموعة التدريب URFD، ثم قام باستخدام شبكة هجينة TD-CNN-LSTM حيث استخدم طريقة التوزيع الزمني Time Distribution مع الشبكة الالتفافية ثم ربط بين الشبكتين CNN و LSTM بشعاع ممدد من طبقة التجميع الأخيرة في شبكة CNN وباستخدام مجموعة التدريب ذاتها، تجاوزت دقة الشبكتين 98%. سلبيات هذه الطريقة عدم الاهتمام بتمييز الوضعيات المرحلية الجزئية المكونة لحدث السقوط، كما أن شعاع الربط بين CNN و LSTM هو شعاع ذو سمات عمومية فقط ولا يمثل السمات التفصيلية.

قام Li وآخرون [110] باقتراح طريقة تعلم غير خاضعة للإشراف لتخفيض زمن تعليم الشبكة، حيث استخدم شبكة التفافية موسعة Dilation CNN (DCNN) لزيادة مجال إدراك الشبكة للصور ذات العمق الملونة،

ثم استخدم شبكة LSTM لكشف الترابط الزمني بين المخارج المتعاقبة لشبكة DCNN، بعدها قام بعمل شبكة مشفر عكسية، حيث تقوم هذه الشبكة بإعادة بناء الإطارات، وحسب منها بطريقة جديدة درجة السقوط التي منها يقدر أن السقوط في المشهد المعاد بناؤه حدث أم لا، أعطت الشبكة دقة 97.1%، سلبية هذه الطريقة أن الشعاع الناتج عن شبكة CNN شعاع يحمل سمات عمومية ولا يهتم بالتصنيف المرحلي الجزئي لحدث السقوط بذاته. قام Zhu وزميله [111] بالدمج بين شبكة YOLO وشبكة OpenPose وذلك للكشف عن الشخص المتزلج على الجليد وكشف حالة السقوط، حيث تقوم شبكة YOLO بالكشف عن الأشخاص في المشهد، بينما تقوم شبكة OpenPose بكشف نقاط الهيكل العظمي للشخص المكتشف، وقد استخدم في ذلك فيديو مسجلة لأناس يتزلجون على الجليد، وقد أعطى الدمج بين الشبكتين دقة 94%، سلبية هذه الطريقة عدم الاهتمام بالسمات الحركية التي تميز السقوط، كما لم يميز محاور الارتكاز أثناء السقوط.

قامت Chen وآخرون [112] بتطوير شبكة YOLO V5، حيث طورت الجزء المتعلق بوحدة الباقي Res Unit، واستخدمت الشبكة المطورة في كشف سقوط كبار السن، وقد أعطت الشبكة دقة 97.2%، بزيادة مقدارها 3.5% عن شبكة YOLO V5 الأساسية. سلبية الطريقة عدم الاهتمام بالسمات الحركية والاكتفاء بالسمات المكانية.

قام Kamel وزميله [113] باقتراح شبكة CNN مؤلفة من ثلاثة طبقات التفاضلية تليها طبقة كاملة الاتصال لكشف حدث السقوط، وقد استخدم صورة MHI كدخل لتلك الشبكة ومجموعة التدريب FDD، وقد أعطت الشبكة المقترحة دقة 95.61% وحساسية 69.17% ونوعية 97.98%، سلبية الطريقة تكمن في كون صور MHI تميز الحركة فقط واتجاهها ولا تميز الوضعيات المرحلية العابرة الساكنة أو محاور الارتكاز أثناء السقوط.

قام Espinosa وآخرون [33] بدمج صورتين التدفق البصري الناتجتين من كاميرتين مستقلتين، ثم استخدم شبكة التفاضلية قام باقتراحها مؤلفة من ثلاث طبقات التفاضلية لاستخلاص السمات الخاصة بالسقوط، ثم تصنيف الحدث كسقوط أو عدم سقوط، أعطت الطريقة دقة 95.64% وحساسية 97.95% ونوعية 83.08%،

سلبية الطريقة الاهتمام بالسمات الحركية فقط وإهمال الوضعيات المرحلية العابرة الساكنة ومحاور الارتكاز، كما أن صور التدفق البصري ذات تعقيد حسابي كبير، ومناسبة للحركات البطيئة وليس للسقوط.

قام Lie وآخرون [34] باستخدام صورة MHI كدخل لشبكة مؤلفة من شبكة CNN تليها شبكة LSTM لكشف حالة السقوط حيث تعمل الشبكة الالتفافية على استخلاص السمات المكانية والزمنية من صورة MHI بينما تقوم LSTM على مراقبة التتابع الزمني لتلك الصور وذلك لكشف السقوط، وصلت دقة الشبكة إلى 97.66%، سلبيات هذه الطريقة تكمن في صور MHI التي تكون قاصرة عن تمثيل الوضعيات المرحلية الساكنة العابرة وكذلك محاور الارتكاز أثناء السقوط، إضافة إلى ذلك فإن الشعاع الناتج عن شبكة CNN شعاع الطبقة ما قبل الأخيرة الذي يحمل سمات عمومية وليست تفصيلية ولا يمثل المراحل الجزئية المكونة للنشاط.

قام Kong وآخرون [36] باقتراح طريقة تحوي ثلاث شبكات التفافية من نوع Resnet-18 على التفرع؛ دخل الشبكة الأولى هي صورة هيئة الشخص (Silhouette Image (SI)، أما دخل الشبكة الثانية فهو صورة MHI لهيئة الشخص، أما الصورة الثالثة فهي الصورة الديناميكية (Dynamic Image (DI التي تميز تاريخ حركة الشخص أيضاً ولكنها تختلف عن MHI في كونها تعطي قيمة إضاءة أعلى للإطارات الأولى وتضمحل في الإطارات اللاحقة، على عكس صورة MHI، وصلت حساسية شبكته إلى 98.8% والنوعية 98%، سلبيات هذه الطريقة تكمن في استخدام عدة صور بصيغ مختلفة لتمثيل الحركة، واستخدام شبكة التفافية من أجل كل صورة من تلك الصور وهذا ما يزيد التعقيد الحسابي ويستهلك موارد الحاسوب.

قامت Khraief وآخرون [35] باقتراح طريقة لكشف السقوط تعتمد على استخلاص صورتين ثنائيتي البعد، إحدهما من التدفق البصري تحوي المطال والاتجاه لعناصر صورة هيئة الشخص (محولة إلى صورة ملونة حيث المطال يقابل اللوحة بينما الاتجاه أو الزاوية تقابل الإشباع)، أما الأخرى فمستخلصة من صورة BMI، بعدها أدخلت الصورتين إلى شبكتين VGG-16 مستقلتين لتعطي خرجين، وقامت بالدمج بينهما بحيث كونا شعاعاً واحداً أدخلته إلى شبكة عصبونية كاملة الاتصال (Fully Connected network(FC لتعطي بعدها التصنيف النهائي للسقوط، حيث أعطت الطريقة حساسية 100% ونوعية 92.5%، ثم طورت فيما بعد العمل

إلى أربعة شبكات [11]، حيث أدخلت شبكتين إنثافيتين إضافيتين، الأولى تأخذ صورة RGB-D، أما الشبكة الثانية فتعتمد على الإزاحة المشتقة من صورة التدفق البصري، حيث أعطت طريقتهما الأخيرة حساسية 100% و نوعية 95.5%، لكن طريقتهما تستهلك موارد الذاكرة لكثرة الشبكات العميقة المستخدمة.

قام Li وآخرون [9] باقتراح طريقة تستخلص هيئة الشخص من الخلفية باستخدام شبكة عصبونية إنثافية، حيث تستخلص سمات عامة وخاصة تعطي خرائط بروز Saliency Maps، ثم تدمج بين تلك السمات لتكون شعاع دخل يمرر إلى شبكة عصبونية عميقة كاملة الاتصال لكشف السقوط، أعطت دقة 98.92%، سلبية هذه الطريقة عدم تمثيل الحركة.

اقترح Yao وزملاؤه [114] طريقة جديدة لكشف السقوط باستخدام شبكة إنثافية سطحية، حيث تعتمد الطريقة على استخلاص هيئة الشخص من المشهد بطريقة نموذج المزيج الغوسي Gaussian Mixture Model (GMM)، ومن ثم تفصل الرأس في تلك الهيئة عن الجسد بواسطة معادلة محددة، بعدها تنشئ قطع ناقص للرأس وقطع ناقص للجسم على عكس الطرق القديمة التي تنشأ قطع ناقص واحد لكامل الهيئة، بعد استنتاج القطعين يستخلص من كل منهما السمات الثلاث التالية: نسبة المحور الصغير إلى الكبير وزاوية الميل والسرعة الشاقولية لمركز القطع الناقص، ويُستحصل في النهاية على ست سمات. بعدها تشكل السمات الستة سمة الحركة (Motion feature) اعتماداً على السلاسل الزمنية، وهذه السمة الحركية مع الزمن هي التي ستمرر كمجموعة معطيات تدريب للشبكة الإنثافية السطحية لكشف خواص الترابط الزمني لشعاع السمة الحركية على طول السلسلة الزمنية، وتتألف الشبكة الإنثافية التي اقترحها من طبقة إنثافية بـ 129 مرشحاً أبعاد كل منها 1x12 يتبعها طبقة Maxpool ثم طبقة كاملة الاتصال يليها تابع Softmax. أعطت هذه الطريقة حساسية مقدارها 90.5%.

قامت Ramirez وآخرون [37] بتصميم نظام لكشف السقوط يعتمد على كشف النقاط الهيكلية للشخص بواسطة شبكة CNN تدعى Alphapose ثم استخدموا تلك النقاط في حساب شعاع سمات مؤلف من 51 عنصر

واستخدموا عدة طرق للتصنيف (RF, KNN, SVM)، أعطت الطريقة نتائج مقاربة بحد وسطي للدقة يساوي 97.65%، يجدر الإشارة هنا إلى أنها استخدمت التعلم العميق فقط لكشف السمات وليس التصنيف.

3-3- الخلاصة

لقد تم في هذا الفصل سرد العديد من الدراسات والطرق في كشف السقوط، والتي تمت في العقد الأخير، وقد صنفت تلك الطرق تبعاً لنوع الحساسات المستخدمة في عملية كشف السقوط إلى: طرق تعتمد الحساسات المحيطية، وطرق تعتمد الحساسات الملاصقة للجسم، وطرق تعتمد على الإبصار الحاسوبي. تقسم طرق الإبصار الحاسوبي من حيث استخلاص السمات الخاصة بالسقوط من خلال الصور إلى طرق هندسية وإحصائية، وطرق تعتمد على تقنيات التعلم العميق، أما بالنسبة لمرحلة التصنيف فتقسم إلى طرق تعتمد على العتبات، وطرق تعتمد على تعليم الآلة؛ فمنهم من اتبع الطرق التقليدية في تعليم الآلة، ومنهم من اتبع التعلم العميق.

لوحظ أن هناك طرق اهتمت بكشف الخواص المكانية لتصنيف السقوط، وأهملت السمات الزمانية أو الحركية، وهذا ما ينعكس سلباً على تصنيف السقوط وتمييزه عن الاستلقاء. وهناك طرق اهتمت بالخواص الزمانية فقط، وأهملت السمات المكانية أو الوضعيات التي يأخذها الجسم أثناء السقوط، وهناك طرق اهتمت بكلا الشكليين من الخواص لكشف السقوط من خلال صور تاريخ الحركة، غير أنها أهملت الوضعيات المرحلية الجزئية التي تكون حدث السقوط وأهملت الوضعيات المرحلية العابرة الساكنة التي يمر بها الجسم أثناء السقوط، كما لوحظ أن الشبكات العميقة متعددة السيول التي تستخدم عدة شبكات بعدة صور لتمثيل السمات الزمانية والمكانية والمراحل الساكنة العابرة تستهلك ذواكر ضخمة ومع ذلك تكون دقة الكشف قريبة من تلك التي هي أقل تعقيداً، كما وجد بأن الشبكات العميقة من النوع الهجين CNN-LSTM التي تعمل على مرحلتين الأولى لكشف الخواص المكانية والثانية لكشف الخواص الزمانية ذات عمق كبير ومقاربة في الدقة من الشبكات العميقة العادية الأقل عمقاً، ففي مرحلتها الأولى لم تميز شبكات CNN-LSTM الوضعيات المرحلية الجزئية المكونة

للنشاط، والشعاع الناتج عنها شعاع عمومي لا يمثل التفاصيل الدقيقة التي تصف المرحلة الجزئية بدقة، حيث أن هذا الشعاع سيكون أداة الربط مع LSTM في المرحلة الثانية. سيتم في الفصل التالي تطوير الشبكة الهجينة CNN-LSTM لتستطيع تمييز الوضعيات المرحلية الجزئية المكونة للنشاط، وتطوير شعاع الربط بين جزئي الشبكة الهجينة. أما الفصل الذي يليه فسيتم فيه طرح صورة حركة جديدة تمثل الوضعيات المرحلية الساكنة العابرة إضافة إلى تمثيل محاور الارتكاز أثناء السقوط التي لم تكن ممثلة في صور الحركة السابقة.

الفصل الرابع

تطوير شبكة CNN-LSTM لكشف السقوط

4-1- تمهيد

إن غالبية الدراسات السابقة تستخدم طرق لطرح الخلفية أو طرق أخرى لاستخلاص هيئة الشخص المتحرك، وهذه العملية تحتاج إلى عناية شديدة في اختيارها، فهي قد تزيل معلومات مهمة في الصورة والتي تفيد في تحديد حالة السقوط، فقد تختفي نتيجة إحدى الطرق مثلاً رجل أو يد أو حتى رأس الشخص مما يؤثر سلباً في السمات والخصائص المستخلصة، وينعكس بالنتيجة سلباً في تمييز السقوط، بل إن حدث السقوط يرتبط بالمكان أو البيئة المحيطة التي حدث فيها السقوط والتي قد يكون لها تأثير مسبب للسقوط، كما أن أغلب الدراسات السابقة لا تميز الوضعيات المرحلية الجزئية المكونة لحدث السقوط بل تميز النشاطات اليومية التي يكون السقوط واحداً منها، لذلك سنطرح في هذا الفصل طريقة جديدة تعتمد على معالجة صورة كاملة (الشخص ضمن البيئة المحيطة) بواسطة التعلم العميق بشبكة عصبونية التلافية CNN، وذلك لكي تكشف الوضعيات المرحلية الجزئية المكونة للسقوط، تليها شبكة عصبونية من نوع ذاكرة طويلة قصيرة-الأمد (Long Short-Term Memory (LSTM) لكشف الترابط الزمني بين تلك الوضعيات المرحلية في سلسلة حدث السقوط، أي سيتم اقتراح تطوير على الشبكة الهجينة CNN-LSTM بحيث تستطيع تميز السقوط بدقة أفضل من تلك الشبكات الهجينة من نوع CNN-LSTM المذكورة في الدراسات السابقة، فمثلاً في الدراسة [34] كان دخل الشبكة هو صور MHI وهذه الصور لا تميز الأجزاء الساكنة من الجسم مما يؤدي إلى ضياع سمات مهمة تصف السقوط، إضافة إلى ذلك فإن الدراسة تميز النشاط اليومي كالسقوط وليس المراحل المكونة للسقوط، وشعاع الربط بين CNN و LSTM شعاع عمومي لا يصف التفاصيل الدقيقة لكل مرحلة.

4-2- بنية الشبكة المقترحة

في هذا الفصل سيتم طرح طريقة جديدة تربط شبكة CNN وشبكة LSTM لتكونا شبكة هجينة من نوع (CNN-LSTM) مع وضع خوارزمية لعمل تلك الشبكة الهجينة بتناسق دون حصول أي خلل فيها عند الكشف عن السقوط وتحديد موسطاتها ومراحل تعليمها.

إن دخل الشبكة الهجينة سيكون صوراً منتقاة من سيل الفيديو المأخوذ من الكاميرا مباشرة أو من ملف فيديو مخزن على أقراص التخزين، وستكون هناك حاجة لتحديد اللحظة الزمنية التي سيبدأ عندها التقاط الصور من الفيديو لمعالجتها وتحليلها، وتحديد هذه اللحظة معتمد على حدوث بداية السقوط، وبداية السقوط هو حركة أو حدث من الأحداث اليومية أو النشاطات اليومية التي يقوم بها الإنسان، وبالتالي غاية الطريقة المقترحة هي تمييز السقوط عن باقي الحركات اليومية التي يقوم بها الإنسان، إذن يجب تمييز حدوث حركة في الفيديو ليبدأ أخذ الصور المعالجة، إن أبسط طريقة لمعرفة حدوث حركة في الفيديو هو مقارنة الصورة الحالية أو الإطار الحالي في الفيديو بالإطار الذي يسبقه، من خلال حساب هيستوغرام الصورة الحالية والتي تسبقها فإن كان هناك فارق أعلى من عتبة δ ، فهذا يعني أنه قد يكون هناك شخص في المشهد، والمشهد عبارة عن أحد الغرف في منزل لشخص كبير السن يعيش فيه وحيداً، وبالتالي إما أن يكون هناك شخص وحيد في المشهد، أو أن المشهد ثابت لا يحتوي شخصاً إنما يحتوي الأثاث (مفروشات وطاولات وكراسي... الخ) حسب الغرفة التي ركبت فيها الكاميرا، ونسمي المشهد الثابت بالخلفية.

بعد اكتشاف الاختلاف في الصور المتعاقبة، يتم البدء بالتقاط الصور واحدة تلو الأخرى، ليتم إدخالها في الشبكة المقترحة، ثم يلي ذلك معالجة الصور بشكل أولي وتخفيض أبعادها بحيث تصبح $227 \times 227 \times 3$ وبذلك تكون مناسبة لإدخالها في شبكة Alexnet الالتفافية.

ثم تُدخل الصورة في شبكة Alexnet المشروحة في الفصل الأول، حيث سيكون دورها في الطريقة المقترحة ليس تصنيف النشاطات اليومية بما فيها السقوط بناءً على سماتها العمومية المعبرة عن مراحلها الجزئية

كما في الدراسات السابقة، ولكن القيام بدل من ذلك بتصنيف المراحل الجزئية المكونة لنشاط السقوط فقط، بينما تصنف أي مرحلة جزئية من النشاطات الأخرى على أنها "آخر"، وذلك تبعاً للمراحل الجزئية التالية المقترحة:

1. "الزلل": هي بداية السقوط التي فيها يفقد الشخص توازنه؛ فقد يتعثر أو يفقد وعيه أو تحصل

له أية حالة تؤدي إلى بداية السقوط.

2. "التساقط": هي مرحلة الانتقال من الاستقامة إلى الاستلقاء على الأرض، وفي هذه المرحلة

تبدأ سرعة الشخص الشاقولية بالتزايد، وهي الفترة الأطول أثناء السقوط.

3. "الاستلقاء": هي المرحلة النهائية للسقوط حيث يصبح الشخص ممدداً على الأرض، إما على

ظهره أو جانبه أو بطنه، وتأخذ أطرافه أوضاعاً مختلفة في هذه الحال أيضاً، ويدل طول الفترة

الزمنية لهذه المرحلة على مدى التأذي الذي تعرض له الشخص.

4. "آخر": هو أية مرحلة جزئية مكونة لأحد النشاطات اليومية ADL غير حركة السقوط.

يلاحظ من الأصناف التي اقترحت أن الشبكة يجب أن تميز الوضعيات المرحلية الجزئية التي يمر

فيها الشخص أثناء السقوط بينما المراحل الجزئية للنشاطات الأخرى تصنف على أنها "آخر"، فهنا خُصص

تميز وضعيات السقوط المختلفة بثلاث حالات ("الزلل" و"التساقط" و"الاستلقاء")، وجُعِلت حالة رابعة ("آخر")

فقط لتمييز المرحلة الجزئية للنشاط اليومي المختلف عن السقوط، أي أصبحت الشبكة بهذا التصنيف أكثر

تخصصاً في تمييز السقوط بمراحله الجزئية وهذا ما سيرفع من دقة الشبكة، بينما الدراسات السابقة كانت تميز

المراحل الجزئية للنشاطات اليومية المراد تصنيفها بما فيها السقوط فكانت بذلك أكثر تعميماً في تمييز تلك

المراحل.

بعد أن تقوم شبكة CNN الالتفافية (Alexnet) بإعطاء تصنيف معين للصورة المعروضة عليها يُؤخذ

خرج تلك الشبكة، كما سيوضح بعد قليل في مخطط خوارزمية العمل، ليكون شعاع دخل لشبكة LSTM التي

ستدرب على سلاسل متتابعة من أشعة الدخل، حيث يمثل كل شعاع وضعاً من الأوضاع التي سيمر بها

الشخص أثناء السقوط، فكلما أعطت شبكة CNN خرجاً اعتماداً على صورة الدخل، أُدخل الخرج كشعاع إلى

خلية LSTM لتقارنه مع الخرج السابق الذي تنبأت به، حيث ستعتمد عليه في إنتاج الخرج الجديد وهكذا حتى تمر مجموعة من الأشعة تحقق بتسلسلها النتيجة النهائية لخلية LSTM، وهذه النتيجة إما أن تكون سقوط (Fallen) أو عدم سقوط (Not_fallen) (أي تميز بين صنفين فقط)، يبين الشكل (4-1) مخطط الشبكة العصبونية الهجينة (Hybrid network) المقترحة لكشف حالة السقوط لدى كبار السن.

يلاحظ من مخطط هذه الخوارزمية أنها تتألف من مرحلتين: المرحلة الأولى هي مرحلة الملاحظة والتخزين للإطارات المتعاقبة في الصّوان (buffer)، فهي تبدأ من مقارنة الإطار السابق بالحالي لمعرفة حدوث أي تغير في المشهد والذي يعني حتماً أن هناك شيئاً ما في المشهد يتحرك وعلى الأغلب شخص لأن المنزل يقطنه مسن وحيد، إن حدث اختلاف تبدأ الخوارزمية بتخزين الإطارات المتتالية في ذلك الصّوان مادام هناك اختلاف.

قُدّر حجم الصّوان بمقدار 120 إطاراً؛ حيث وُجد نتيجة متابعة العديد من الفيديوهات لحالات السقوط بأن أطول فترة لحدوث السقوط هي أربع ثوان (أي قد يتعثر الشخص ثم يحاول التوازن ثم يسقط متأدياً) بينما يتم السقوط الفعلي من الزلزال إلى الاستلقاء في فترة تتراوح بين 300ms و 500ms، وبما أن أغلب كاميرات المراقبة تعمل بمعدل 25f/s فإن زمن الإطار هو 0.04s، وعليه فإن 15 إطاراً تكافئ فترة السقوط الفعلي، بينما السقوط بشكل عام قد يستغرق زمن 120 إطاراً. يقوم الصّوان بمجرد دخول الإطارات إليه بإرسالها واحداً تلو الآخر لمرحلة المعالجة الأولية حيث يخفض حجم الصورة بطريقة الهرم الغوسي (Gaussian pyramid)، فيصبح كل إطار (صورة) بحجم 227x227x3 لتدخل إلى المرحلة الثانية، وهي الشبكة الهجينة المطورة المقترحة Alexnet-LSTM، إذ تدخل الصورة إلى شبكة Alexnet الالتفافية لتقوم بتصنيفها إلى أحد الأصناف الأربعة المقترحة كمراحل جزئية مكونة لنشاط السقوط وهي "الزلزال" أو "التساقط" أو "الاستلقاء" أو "آخر".

إن التعديل الذي أحدث في شبكة Alexnet هو في مرحلة الخرج، إذ أُلغيت الطبقة FC8 ذات 1000 عصبون المعبرة عن ألف صنفٍ دُرِبَت عليها، واستُبدلت بطبقة FC8 ولكن بأربعة عصبونات دلالة على الأصناف الأربعة المقترحة التي ستميز بينها.

أما دخل خلية LSTM فلم يُؤخذ من الطبقة FC8 لأن هذه الطبقة تعطي أحد أربعة تصنيفات، وكل صنف فيها يعبر عن تعميم للعديد من الوضعيات، فمثلاً تصنيف الاستلقاء يمثل العديد من الوضعيات للشخص على الأرض، مستلقياً على ظهره، أو على جنبه، أو على بطنه. ولكل حدث سقوط هناك سلسلة من الوضعيات المرحلية الخاصة الجزئية المترابطة التي تشكل بتتاليها حالة السقوط النهائية، وعليه فلا ينفع أخذ التعميم من الطبقة FC8 ليكون دخلاً لشبكة LSTM.

في الحقيقة إن الدراسات السابقة كانت تستخدم خرج الطبقة ما قبل الأخيرة (مثل FC7 في شبكة Alexnet) كدخل لشبكة LSTM، وقد تم أخذ خرج هذه الطبقة FC7 وتجربتها ولكن نتيجة التدريب لم تعطي دقة جيدة، وذلك لكون الطبقة FC7 تحمل نوعاً من التعميم، لذلك تم اقتراح ما يلي كدخل لشبكة LSTM: أخذ خرج الطبقة العصبونية FC6 التي تحمل شعاع صفات الصنف النهائية التفصيلية المؤلفة من 4096 صفة (حيث في هذا الشعاع يختلف كل "زلل" عن زلل آخر وكل "تساقط" عن تساقط آخر من ناحية تغير قيم تلك الصفات) مع خرج الطبقة FC8 (الذي هو شعاع التعميم النهائي المؤلف من أربعة صفات) ليشكلا شعاع دخل نهائي لشبكة LSTM، أي أن شعاع دخل خلية LSTM هو x_i

$$x_i = [FC6, FC8]$$

وهو مشكل من شعاعين FC6 و FC8 بطول 4100 عنصر، وتدل i على ترتيب الشعاع ضمن إحدى السلاسل X_n المراد تصنيفها، وبهذا تتميز كل سلسلة X_n (حيث n هو رقم السلسلة) بمجموعة من الأشعة المتعاقبة $x_i \in X_n$ ، وبهذا تستطيع خلية LSTM توقع خرج السلسلة النهائي بدقة، الذي هو خرج الشبكة، فإما "سقوط" (Fallen) أو "عدم سقوط" (Not_fallen).

4-3- مرحلة التدريب والاختبار

4-3-1- معطيات التدريب

لقد اعتمد في تدريب الشبكة العصبونية الهجينة على عدة مجموعات صور :

1. مجموعة FDD أو ما تعرف أيضاً Le2i، تتألف هذه المجموعة من 221 مشهداً، تتضمن

مختلف النشاطات اليومية بما فيها السقوط، ومسجلة بمعدل إطار 25f/s وأبعاد 320x240،

والمشاهد السابقة غنية بالخلفيات المتنوعة من مفروشات وكراسي وغيرها، فهناك مشاهد لغرفة

الجلوس والمكتب ... إلخ.

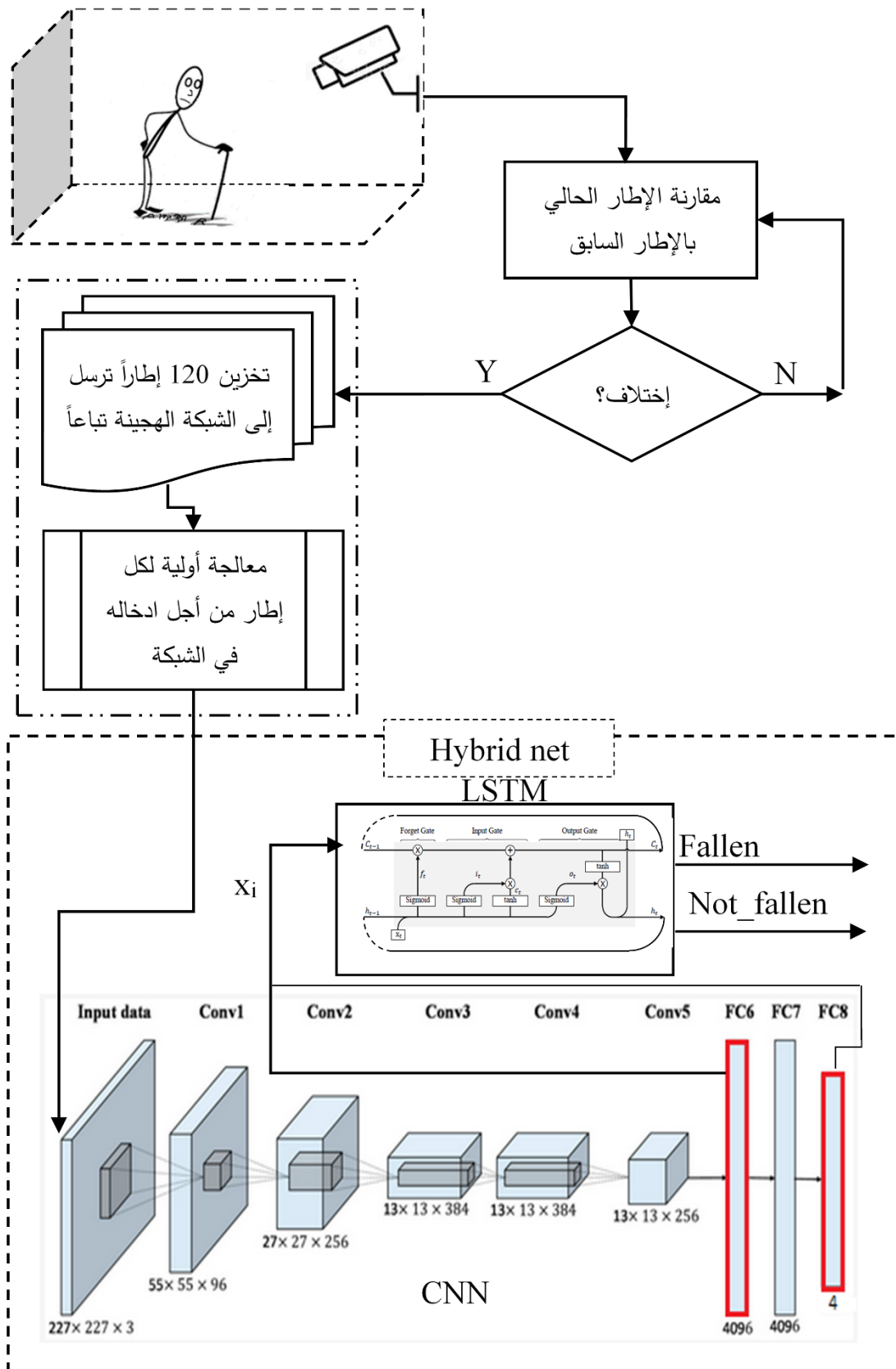
2. مجموعة URFD وتتألف من 30 مشهد سقوط و40 مشهداً للنشاطات اليومية، بأبعاد

.640x480

3. مجموعة U. Montreal تحوي 24 مشهداً، كل مشهد مصور من ثمانية مواضع مختلفة، 22

مشهداً منها تتضمن السقوط، بينما مشهذان يتضمنان النشاطات الحياتية اليومية، سجلت تلك

المشاهد بمعدل إطار 120 f/s وأبعاد 720x480، وجميع تلك المشاهد مسجلة في المختبر.



الشكل 4-1 المخطط التدفقي للخوارزمية والشبكة الهجينة المقترحة

الشكل (4-2) يبين بعض الصور لحالات السقوط من تلك المجموعات، التي غالباً ما تعاني من قيود

ومحددات تؤثر في عملية التدريب وهذه القيود يمكن تلخيصها في البنود التالية:

1. تتكون مجموعة المعطيات الفيديوية عادةً من البيانات المسجلة دون تغيير الظروف (غالباً في

ظروف المختبر، بدلاً من الظروف الواقعية)، وتكون إضاءة المشهد المراد تحليله بالكامل

موحدة وليس كما في الواقع حيث تختلف الإضاءة وأثاث المكان. وهذا ما يلاحظ في مجموعة

.U. Montreal

2. يكون ذات الشخص مستخدماً في التدريب ولربما في الاختبار.

3. الحركات المؤدية للسقوط التي يقوم بها الأشخاص في المشهد مصطنعة، وغالباً ما يكررها

الشخص ذاته، وما يزيد الأمر سوءاً كون التصوير من نفس الزاوية ونفس المكان.

4. في جميع الحالات تقريباً، يتم إجراء السقوط على وسادة مبطنة، والتي غالباً ما يتباين لونها

بشكل واضح مع ملابس الشخص.

للتخلص من القيود السابقة طُبق ما يأتي:

1. تم توحيد أبعاد جميع مجموعات المعطيات بقياس 227x227 وبمعدل إطار 25 f/s، وذلك

لتكوين مجموعة معطيات واحدة.

2. أجريت عملية زيادة المعطيات (Data augmentation) لكيلا تدخل الشبكة المدربة فيما يسمى

الموائمة الزائدة، حيث زدنا عدد مشاهد الفيديو اعتماداً على المشاهد القديمة، وذلك من خلال

عكس اتجاه الشخص في المشهد، وإزاحته عن المكان الذي هو فيه، وتغيير الإضاءة، واقتطاع

أجزاء من المشهد.

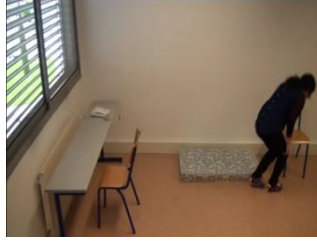
3. أعيد فرز المشاهد بحيث أعطيت المشاهد الممثلة للسقوط سمة السقوط، بينما الأخرى اعتبرت

نشاط حياتي يومي.

4. شُكلت مجموعة صور تدريب من المجموعات الثلاثة كي تتدرب الشبكة على شكل أكثر من

شخص، وبالتالي يكون هناك نوع من التعميم على هيئة الشخص.

بعد ذلك أُخذ من المشهد الحاوي على السقوط لقطة السقوط، وجُعِلت اللقطة الواحدة مساوية لأربعة ثواني، وهذا الزمن موافق لحدث السقوط المؤذي الذي نوقش في الفقرة السابقة، طبعاً وهذا ينطبق على المشاهد الخالية من السقوط، وبعد الحصول على اللقطات، حُولت إلى صور، وأعطيت كل صورة تسمية (Label) من التسميات الأربعة للأصناف المراد تدريب شبكة Alexnet عليها، وخزّنت الصور الخاصة باللقطة في مجلد مستقل يحمل رقماً n يميز اللقطة P_n ، وبهذا تكون مجموعة المعطيات P هي $P = \{P_1, P_2, \dots, P_n\}$ جاهزة لعملية تدريب الجزء الأول من الشبكة الهجينة وهي شبكة Alexnet.



صورة انحناء فتاة من مجموعة FDD



صورة سقوط رجل من مجموعة URFD



صورة زل شخص وبداية سقوطه على الأرض من مجموعة U. Montreal

الشكل 4-2 صور متنوعة من مجموعات معطيات التدريب.

4-3-2- مرحلة التدريب

سيتم في هذه المرحلة تدريب الشبكة الهجينة (CNN-LSTM) على تمييز صنفين إما سقوط أو عدم سقوط، ولكن بدل تدريب كامل الشبكة الهجينة في مرحلة واحدة، سيجزئ التدريب على مرحلتين: مرحلة لتدريب شبكة Alexnet ومرحلة لتدريب LSTM، ومع أن هذا سيستغرق وقتاً أطول إلا أن له مزيّتين، الأولى هو جعل كل شبكة خبيرة في مجالها، والثانية وهي الأهم، ألا وهي التخلص من العمق الكبير للشبكة الهجينة أثناء التدريب

والذي ينعكس بدوره على تلاشي التدرج، حيث يكون له أثر سيء في طور الانتشار العكسي أثناء التعلم، مما ينتج عنه قلة فعالية ودقة التصنيف النهائي.

4-3-2-1- المرحلة الأولى من التدريب

في هذه المرحلة تم حذف الطبقة FC8 التي تميز بين 1000 صنف، وأضيفت طبقة FC8 مؤلفة من أربعة أصناف فقط وهي الأصناف التي ستميزها الشبكة المقترحة، وبعدها دربت شبكة Alexnet المعدلة بنقل التعلم على تمييز الأصناف الأربعة المقترحة، باستخدام الماتلاب 2020a وذلك بعد تنصيب قالب أداة التعلم العميق الداعمة لحزمة Alexnet (Deep Learning Toolbox Model for Alexnet)، وقد قسمت مجموعة معطيات التدريب P إلى 70% للتدريب و 15% للاختبار و 15% للصلاحية (Validation)، أما بالنسبة لمعدل التعليم (Learning rate) فقد جُعل بمقدار 0.001، وذلك لكي يتم تعديل الأوزان أثناء التدريب بشكل صغير في كل مرة، أما بالنسبة لعدد دورات التدريب (Epochs) فقد اختيرت مساوية 40 دورة، أما حجم الدفعة (Batch size) من الصور فهو 64.

4-3-2-2- المرحلة الثانية من التدريب

بعد أن تم تدريب شبكة Alexnet المعدلة على تصنيف كل صورة في أحد الوضعيات الأربعة المقترحة، مررت كل صورة pi من مجموعة التدريب على Alexnet المعدلة وحُزن خرج الطبقة FC6 وكذلك خرج الطبقة FC8 كشعاع xi موافق لتلك الصورة pi، فتم الحصول في نهاية العملية على مجموعة سلاسل الأشعة X

$$X = \{X_1, X_2, \dots, X_n\}$$

حيث رتبت في مجلدات مشابهة تماماً لمجلدات الصور التي رتبت سابقاً، فمثلاً السلسلة X1 تحتوي على مجموعة أشعة كل منها $x_i \in X_1$ موافقة للمجموعة P1، وبذلك يكون قد حُصل على مجموعة التدريب المتسلسلة الخاصة بتدريب شبكة LSTM، كما تم توليد مجموعة الهدف T كملف من نوع إكسل أعمدته كل منها يحمل نفس رقم السلسلة الموافقة، وكل عامود عبارة عن شعاع فيه قيمتان أحدهما للسقوط والأخرى لعدم

السقوط، ثم تم تدريب شبكة LSTM على المجموعة $\{X, T\}$ بواسطة أداة التعلم العميق ذاتها، بحيث تميز في النهاية بين نشاط السقوط و نشاط عدم السقوط.

4-4- النتائج والمناقشة

أعطت نتائج الاختبار على شبكة Alexnet المعدلة التي تم تدريبها على مجموعة التدريب P دقة وصلت إلى 98.46%، بينما أعطت شبكة LSTM التي دربت على المجموعة X دقة وصلت إلى 98.6%، وذلك لأن كل شعاع $x_i \in X$ يتكون من صفات أكثر تفصيلاً وهو خرج الطبقة FC6، وصفات أكثر تعميماً وهي خرج الطبقة FC8، فكانت النتيجة جيدة جداً، مقارنة بمرحلة التدريب التي دربت فيها الشبكة LSTM على أشعة ناتجة فقط عن الطبقة FC7، حيث أعطت دقة 97.15% حينها. وبناءً على ما سبق تكون دقة الشبكة الهجينة المقترحة هي 98.6%.

يبين الشكل (3-4) الدقة مقابل عدد مرات التدريب لشبكة Alexnet، أما الشكل (4-4) فيبين الدقة لشبكة LSTM. يلاحظ من الشكل (3-4) أن شبكة Alexnet قد تعلمت بنقل التعلم التعرف على الأصناف الأربعة خلال أربعين دورة تدريب، ولكن الدقة الأعلى المساوية 98.469% حصلت عليها الشبكة في دورة التعلم 13 أي أنها خلال عدد قليل من دورات التعلم وصلت إلى دقة عالية، ويعود الفضل في ذلك إلى عدة عوامل:

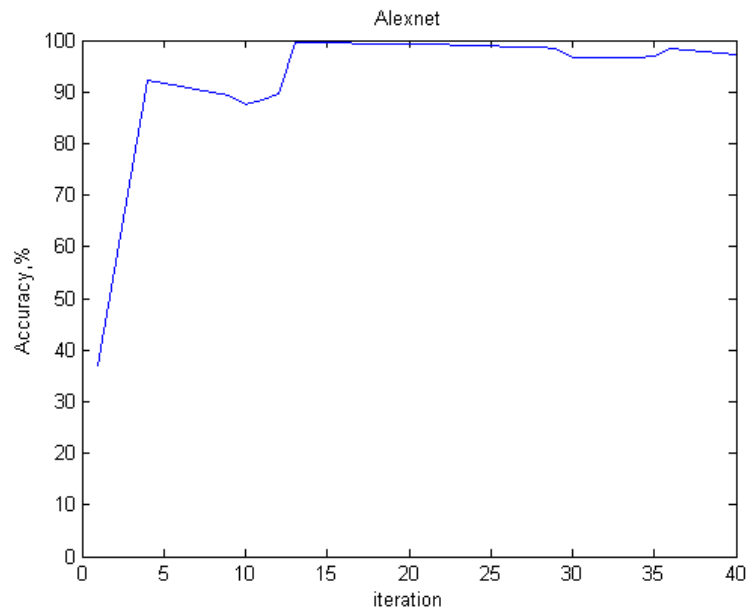
- طريقة نقل التعلم التي تحافظ على ما تعلمته الشبكة سابقاً وتتطوّر منه لدورة تعلم جديدة بمعطيات تعلم جديدة، ويفيد هذا النوع من التعلم في تجاوز مشكلة قلة أمثلة التعلم الجديدة.

- تدريب الشبكة على كم كبير من مجموعات معطيات التعلم (ثلاثة مجموعات)، مضافاً إليها صور جديدة لتلك المجموعات من خلال عملية زيادة المعطيات.

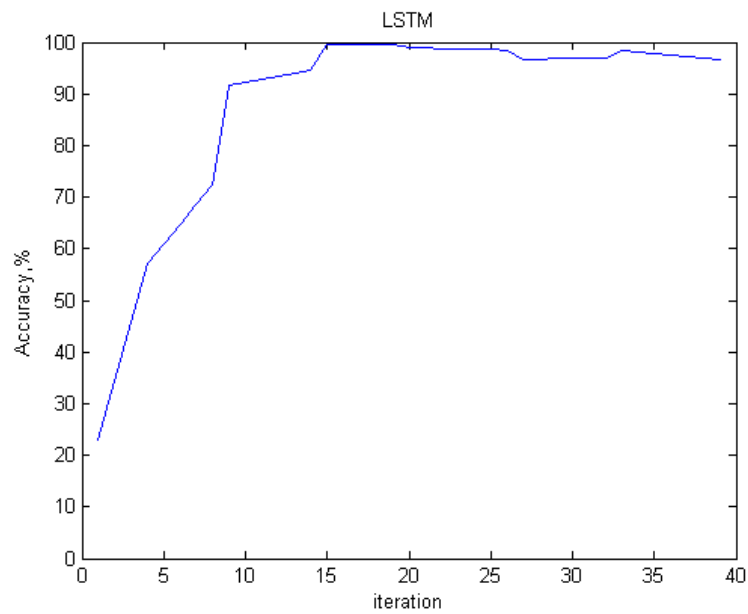
- استخدام كامل الصورة في التعلم؛ أي أننا نستخدم الشخص (الصورة الأمامية)، والغرفة مع محتوياتها (الصورة الخلفية)، حيث تبين نتيجة تعليم الشبكة أن اقتطاع صورة الشخص من المشهد والتعليم عليها ليس مفيداً في حالة الحركة التفاعلية بين الشخص والمكان كما في مثال السقوط، لأن السقوط مرتبط بأشياء

حول الشخص قد تكون سبباً في سقوطه؛ أي قد يكون هناك رابط بين سقوط الشخص وخلفية المشهد، مثل تعثره بشيء على الأرض، لذلك يجب أن تُعلم الشبكة على كامل المشهد، وهذا الطرح يتوافق مع النتيجة التي توصلت لها Sherin P [30].

يلاحظ أن هذه الدقة التي تم الحصول عليها أعلى من الدقة في كل من [29, 31, 33, 107, 108].
 يبين الشكل (4-4) الدقة مقابل دورات التدريب الأربعين التي دربت عليها الشبكة، حيث يلاحظ أن الدقة الأعلى التي وصلت إليها شبكة LSTM وبالتالي الشبكة الهجينة المقترحة هي 98.6% في دورة التدريب 15؛ أي خلال عدد قليل من دورات التدريب تم الوصول إلى دقة تصنيف عالية لتصنيف الحركة بين السقوط أو عدمه، ويعود السبب في ذلك إلى التركيبة الجديدة لشعاع الدخل $xi=[FC6,FC8]$ المكون من صفات تفصيلية يمثلها خرج الطبقة FC6 من شبكة Alexnet، وصفات عامة أو تصنيف عام يمثلها خرج الطبقة FC8 من شبكة Alexnet. يلاحظ أن هذه الدقة التي تم الحصول عليها أعلى من الدقة في حالة الشبكة المقترحة من قبل Lie [34] التي بلغت 97.66% فقط، وأعلى من الدقة لشبكة Sangwan [32] التي بلغت 91.62%، وكذلك أكبر من الدقة لشبكة Salimi [109] التي بلغت 98.09%، وأيضاً أكبر من الدقة لشبكة Li [110] التي بلغت 97.1%، وجميع تلك الطرق في هذه الدراسات السابقة تعتمد الشبكة الهجينة CNN-LSTM في التصنيف.



الشكل 3-4 منحنى الدقة لشبكة Alexnet المعدلة مقابل عدد دورات التدريب



الشكل 4-4 منحنى الدقة لخلية LSTM مقابل عدد دورات التدريب

كما يبين الجدول (1-4) مصفوفة الالتباس (confusion Matrix) لشبكة Alexnet.

الجدول 4-1 مصفوفة الالتباس Confusion Matrix

الوضع المرحلي الحقيقي	عدد صور الاختبار	الوضع المرحلي المُتنبأ به			
		زلل	تساقط	استلقاء	آخر
زلل	315	313	2	0	0
تساقط	945	5	936	0	4
استلقاء	315	0	0	312	3
آخر	1485	13	10	10	1452

يمكن أن نحسب من الجدول (4-1) الحساسية والنوعية لكل من الوضعيات المرحلية المختلفة المقترحة كما في الجدول (4-2).

الجدول 4-2 الحساسية والنوعية لكل من الوضعيات المرحلية.

الوضع المرحلي	الحساسية Sensitivity	النوعية Specificity
زلل	99.36%	99.34%
تساقط	99.04%	99.43%
استلقاء	99.04%	99.63%
آخر	97.77%	99.5%

يلاحظ من الجدول (2-4) بأن الحساسية والنوعية عالية لكل من الوضعيات الأربعة، وهذا ما يؤكد صوابية اختيار تمييز الوضعيات المرحلية الجزئية المكونة للنشاط عند العمل على اكتشاف حركات الجسم البشري مثل نشاط السقوط، إذ بجعل التركيز على المراحل الجزئية للنشاط المطلوب فقط تزداد حساسية الشبكة، كما يلاحظ انخفاض بسيط في حساسية الوضعية "آخر" كونها تحتوي على وضعيات تشبه بعض الشيء الوضعيات المرحلية الجزئية المكونة للسقوط.

يبين الجدول (3-4) مصفوفة الالتباس (confusion Matrix) لشبكة LSTM.

الجدول 3-4 مصفوفة الالتباس Confusion Matrix لشبكة LSTM.

الوضع الحقيقي	عدد سلاسل الاختبار	الوضع المتنبأ به	
		سقوط	عدم سقوط
السقوط	35	35	0
عدم السقوط	37	1	36

ويمكن أن يحسب من الجدول (3-4) الدقة والحساسية والنوعية لوضعية السقوط كما في الجدول (4-4).

الجدول 4-4 الحساسية والنوعية لشبكة LSTM.

النشاط	الدقة Accuracy	الحساسية Sensitivity	النوعية Specificity
السقوط	98.6%	100%	97.29%

يلاحظ من الجدول (4-4) بأن الحساسية لحالة السقوط في شبكة LSTM هي 100%، أي أن شبكة LSTM قادرة على تمييز حالة السقوط بشكل كامل. أما بالنسبة للحساسية في حال عدم السقوط فإنها 97.29%. يعود سبب ازدياد هذه الحساسية إلى طبيعة شعاع الدخل الجديد [FC6,FC8] المستخدم في شبكة LSTM والمشتق من خرج طبقة عمومية FC8 وخرج طبقة تفصيلية FC6 في شبكة Alexnet.

4-5- الخلاصة

في هذا الفصل تم طرح طريقة جديدة لكشف أحد الحركات التي قد يكون لها أثر سلبي على صحة كبار السن أو حتى قد تشكل خطراً على حياتهم، وذلك من خلال تطوير الشبكة الهجينة (CNN-LSTM)، التي تدمج بين شبكة CNN ممثلة بشبكة Alexnet بعد تعديلها وشبكة LSTM، وكان التطوير في اقتراح شعاع سمات جديد من شبكة Alexnet يجمع بين الصفات العامة والتفصيلية لكي يستخدم كدخل لخلية LSTM، واستخدمت طريقة نقل التعلم في تعليم شبكة Alexnet المعدلة، مع معالجة كامل الصورة (الشخص والخلفية) ضمن الشبكة.

أما التطوير الآخر فهو في جعل شبكة Alexnet تميز الوضعيات المرحلية الجزئية المكونة لحدث السقوط فقط وهي: الزلل، والتساقط، والسقوط، أما الوضعيات المرحلية الجزئية المتعلقة بالنشاطات الأخرى فاقترح لها تصنيف "آخر" يدل على أن الوضعية ليست من تلك الوضعيات المتعلقة بالسقوط بل تتبع لأحد النشاطات الحياتية اليومية الأخرى، وهذا ما لم يكن مطروحاً في الدراسات السابقة التي كانت تهتم بكشف النشاط بشكل كامل وليس أجزأؤه، وتعطي ذات الخصوصية لكشف النشاطات المراد كشفها بما فيها السقوط.

وبهذا تميز شبكة Alexnet السمات الحيزية في الصور المتعلقة بالمراحل الجزئية المكونة للسقوط، أما شبكة LSTM فدربت على الشعاع المستخلص من Alexnet لتمييز بعد مرور مجموعة متسلسلة من تلك الأشعة بين حالة السقوط أو عدم السقوط، أي أن LSTM تميز السمات الزمنية التي تربط بين السمات المكانية المتعاقبة. تم تدريب الشبكة الهجينة المطورة على مرحلتين، إذ أعطت الطريقة دقة عالية جداً، وتميز كامل لحالة السقوط. إلا أن أخذ الصورة كاملة للشخص مع الخلفية مع ما تحمله من معلومات مهمة تعتبر بالنسبة للشخص الذي تركب عنده المنظومة نوعاً من انتهاك الخصوصية، لذلك يجب البحث عن صورة تعبر عن الشخص دون معرفة التفاصيل الدقيقة له، ولكنها تحمل سمات حيزية بما فيها الوضعيات المرحلية الساكنة

والمراحل الجزئية للسقوط وتحمل سمات زمانية يمكن من خلالها الاستغناء عن LSTM، بحيث تمرر تلك الصورة على شبكة التلافية واحدة فنحصل على التصنيف النهائي.

الفصل الخامس

استخدام صورة الحركة STTFI المقترحة في منظومة كشف السقوط

5-1- تمهيد

يلاحظ من الدراسات السابقة في الفصل الثالث، أنها انتهجت أربعة طرق من ناحية كشف السقوط بتقنيات التعلم العميق؛ الأولى استخدام صورة واحدة لكشف السقوط من خلال تحديد وضعية الاستلقاء على الأرض، وهذا لا يعني بالضرورة السقوط، أما الطريقة الثانية فهي استخدام صورة تشتمل بمحتواها على تغيرات حركة الشخص في عدة إطارات frames من الفيديو مثل التدفق البصري، أو أن تشتمل على التغيرات المكانية والزمانية مع كمية الحركة مثل MHI، أما النهج الثالث فهو يستخدم تعدد السيول multi-streams ويكون لكل سيل شبكة التقافية ودخل مختلف، وهذه الطريقة تستهلك موارد الذاكرة في الحاسوب وتتطلب حسابات طويلة، والدقة التي تقدمها لا تزيد كثيرا عن الطرق السابقة، أما النهج الرابع فيعتمد شبكة عميقة مؤلفة من مرحلتين المرحلة الأولى شبكة التقافية تكشف السمات المكانية في الصورة، والثانية شبكة LSTM لكشف الترابط الزمني بين تلك السمات لتقرر النتيجة النهائية، وكذلك يعتبر هذا النهج كسالفه من حيث استهلاك موارد الحاسوب، ويبقى النهج الثاني هو الأفضل، وخاصة صور MHI ولكن المشكلة في أن تلك الصور مع أنها تعبر عن تغيرات كمية الحركة وسيرها مع الزمن إلا أنها أغفلت أن الإنسان كائن هيكلي مفصلي [115] وليس مكعباً متحركاً أو كرة صماء تتحرك كل نقاطها معاً، فأثناء السقوط أو أي حركة أخرى هناك أجزاء من الجسم أو مفاصله تبقى ساكنة ومنها ما يكون محاور ارتكاز، واستخدام التدفق البصري أو MHI بأشكاله المختلفة سيحذف من تلك الصور العناصر الثابتة والتي تمثل أوضاع مرحلية ذات شكل معين تساهم في إعطاء خصائص مكانية وزمانية تفيد في عملية التصنيف النهائي، وهذا ما لوحظ في الطريقة المقترحة في الفصل الرابع حيث بالاعتماد على كشف الأوضاع المرحلية الجزئية المكونة للنشاط زادت دقة كشف حالة السقوط، لذلك سيتم طرح طريقة جديدة تتلافى سيئة MHI والطرق المطورة منها؛ المتمثلة في عدم إظهار الأجزاء الساكنة من هيئة

الشخص التي تحمل سمات الشكل والمكان والأزمنة المرتبطة بها، كما تتلافى سلبية التغطية الذاتية خاصة عندما يكون محور الحركة موازياً لمحور الرؤية، وتستطيع الإجابة عن كيف ومتى تمت الحركة أو النشاط.

5-2- صورة أمامية مكدسة بعلامات زمنية Stacked Time-Tagged Foreground Image (STTFI)

بما أن الجسم البشري أثناء الحركة الطبيعية أو غير الطبيعية كالسقوط تكون فيه أجزاء ساكنة كمفاصل الأطراف التي تعتبر محاور ارتكاز، أو أن الجسم أثناء السقوط يمر بمراحل عابرة لحظية يكون فيها الجسم في أغلبه ساكناً وأخذاً شكلاً معيناً، فإن تلك الأجزاء ستحتوي على سمات بصرية مكانية وزمانية يكون لها أثر كبير في تحديد نوعية الحركة؛ لذلك يجب أن تتضمن أي صورة حركة مشتقة من عدة إطارات هذه الأجزاء موسومة باللحظة الزمنية التي حدثت فيها، فإهمال هذه السمات سيؤثر سلباً على تصنيف الحركة وبالتالي على دقة تصنيف النظام المقترح.

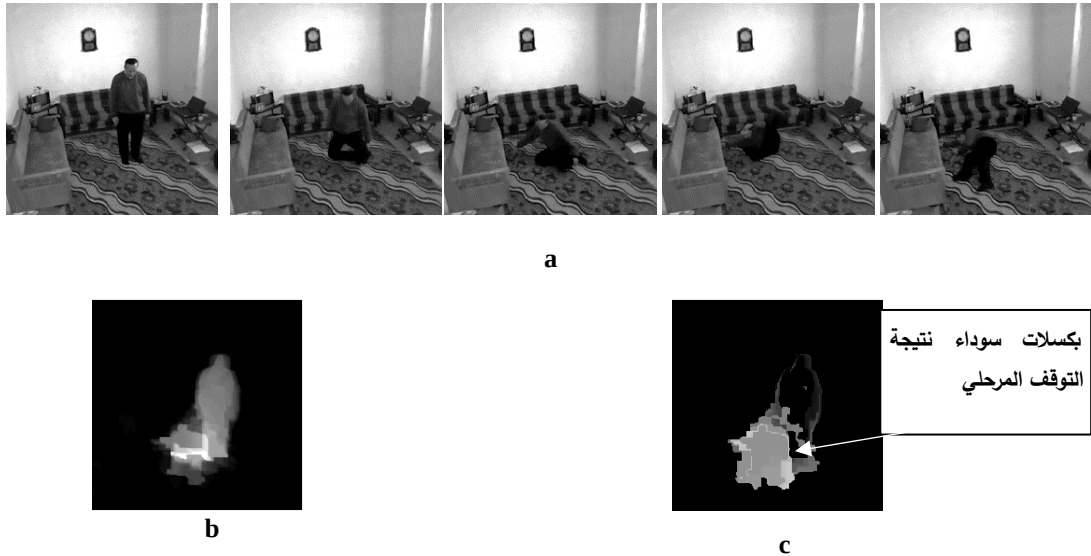
لحل هذه الإشكالية اقترحت صورة حركة جديدة تدعى صورة أمامية مكدسة بعلامات زمنية Stacked Time-Tagged Foreground Image (STTFI) وتحسب هذه الصورة وفق المعادلة التالية:

$$STTFI(x, y, t) = \begin{cases} STTFI(x, y, t - 1) + KT + 1 & \text{if } f(x, y, t) = 1 \\ STTFI(x, y, t - 1) & \text{otherwise} \end{cases} \quad (1-5)$$

حيث تمثل $f(x, y, t)$ أحد بكسلات صور هيئة الشخص في اللحظة t وهي إما صفر أو واحد لأن صورة الهيئة ثنائية، أما K فهو رقم الإطار المختار من الفيديو، بينما T عبارة عن جزء عشري ثابت يختار بناء على عدد الإطارات الأعظمية المشكلة للصورة STTFI، فمثلاً لو كان عدد الإطارات الأعظمي هو N إطاراً وزمن الاعتيان T_s فإن T تختار بحيث $T \ll \frac{T_s}{N}$ حتى لا يتداخل الجزء الصحيح مع العشري. وبالتالي ستكون قيمة كل عنصر من الصورة الجديدة مؤلف من جزأين جزء صحيح وجزء عشري؛ الجزء العشري يعبر عن الوسوم الزمنية التي يمر بها عنصر الصورة في الموضع (x, y) ، بينما الجزء الصحيح فيعبر عن القيمة التراكمية

للبكسل في الإحداثي (x,y) ، وعليه تكون الصورة STTFI عبارة عن هيئة الشخص المتعاقبة المتراكمة على شكل بكسلات قيمها تعبر عن تكاثف بكسلات الهيئة، فكلما كانت هيئة الشخص أكثر ثباتاً في المكان كانت تلك القيمة للبكسلات المكونة لهيئة الشخص أعلى قيمة (أكثر إضاءة)، يمثلها الجزء الصحيح بينما الأجزاء العشرية تعبر عن الأزمنة التي مرت على تلك البكسلات حتى أخذت تلك القيم، في حين يعبر امتداد الهيئة على مساحة واسعة من الصورة عن الحركة لتلك الهيئة والقيم المتدنية للبكسلات تعبر عن مدى سرعة تلك الحركة.

يمكن توضيح ذلك من خلال تكوين صورة STTFI لوضعية السقوط لشخص من مجموعة معطيات التدريب التي تم تصويرها -لغرض اقتراح وتدريب الطريقة الجديدة- كما في الشكل (5-1).



الشكل 5-1- a مجموعة إطارات لحركة السقوط في حال فقدان الوعي، b- صورة STTFI المستنتجة منها، c- صورة

$$MHI(\tau=40)$$

يمثل الشكل (5-1- a) مجموعة من الإطارات مأخوذة من مشهد لسقوط شخص مغماً عليه من مجموعة معطيات التدريب المنشأة في هذا البحث FD22، أما الشكل (5-1- b) فيبين صورة STTFI، بينما الشكل (5-1- c) فيبين صورة MHI التي يلاحظ منها أن الوضعيات التي يكون فيها الشخص في حالة توقف مرحلي عند

اصطدام ركبتيه بالأرض وحين اصطدام إليته بالأرض، تكون البكسلات المقابلة لها سوداء وذلك لعدم وجود حركة، ونجد ذلك أيضاً في حالة الوقوف قبل السقوط حيث يرى فقط كفاف محيط به بينما جسمه الساكن تكون بكسلاته سوداء أيضاً، وهذا يؤدي إلى فقدان السمات المكانية المتعلقة بالشكل والتي تلعب دوراً في تحديد طبيعة الحركة، بينما يبين تدرج الرماديات اتجاهه الحركة، ويلاحظ بأن الحركات الحديثة قد غطت أو مسحت مسار الحركات القديمة.

بينما يلاحظ في الصورة الجديدة STTFI الشكل (b-1-5) بأن الوضعيات المرحلية الساكنة واضحة تماماً، حيث تعتبر الساق اليمنى الأفقية محور ارتكاز ودوران أثناء سقوط الشخص ويمثلها الخط المستقيم الأفقي ذو السطوع الكبير، وحتى في حال استقرار جسم الشخص مستلقياً على الأرض يلاحظ أن بكسلات وضعية الاستلقاء الأخيرة تماثل تماماً وضعية الاستلقاء الأخيرة في الشكل (a-1-5) الصورة الأخيرة على اليمين، كما أنه في حالة الحركة تكون قيمة البكسلات متناسبة عكساً مع الحركة؛ فكلما كانت الحركة سريعة كانت قيمة البكسل أقل قيمة، وعلى عكس ذلك فالوضعيات شبه الساكنة أو الساكنة تكون قيمة البكسل كبيرة جداً، ويجب التنويه أنه في حال البكسلات المتحركة أو الساكنة فإن هناك جزء عشري يلعب دوراً في تحديد الأزمنة التي كان فيها البكسل يحمل تلك القيم. أما في MHI فقيمة البكسل المتحرك تعبر عن مدى حدثه فالحديث ذو شدة عالية والقديم ذو شدة منخفضة تتخفف مع قدوم إطارات جديدة لتشكيل صورة MHI، كما أن البكسلات الحديثة تغطي على القديمة التي تقع في ذات موضعها، وهذا ما تم تلافيه في صور STTFI.

يمكن هنا ذكر الفروقات بين صور STTFI المقترحة وصور MHI:

1. حتى يتم تطبيق التعديل على قيمة البكسل في المعادلة يجب:

 - أن يكون - بالنسبة STTFI - البكسل في الموضع (x,y) من الإطار الحالي، منتمي لهيئة الشخص.
 - أن يكون - بالنسبة MHI - الفرق بين قيمتي ذات البكسل في إطارين متعاقبين يمثل الشدة العظمى للبكسل بالقيمة المطلقة.

2. من حيث التعديل:

- يتم التعديل في STTFI بزيادة واحد إلى القيمة السابقة مضافاً له دلالة زمنية على الإطار الذي وجد فيه الواحد.

- يتم التعديل في MHI بحذف القيمة السابقة وجعل القيمة الجديدة تساوي ثابت τ اختياري بين 1 و255.

3. في حال عدم التعديل:

- تبقى قيمة STTFI السابقة على حالها

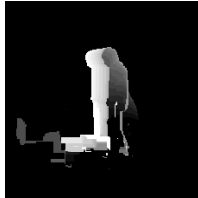
- تنقص قيمة MHI السابقة بمقدار واحد وتقارن مع الصفر لتحذف إن قلت عنه.

تبين الأمثلة التالية المطبقة على مجموعة صور من مجموعة التدريب FD22 الفروقات بين تمثيل صور STTFI للنشاط وبين تمثيل صور MHI للنشاط، وذلك من أجل قيم مختلفة لـ τ في حال صور MHI.

يبين الشكل (5-2) لقطة فيديو لشخص انحنى ثم عاد واعتدل وبدء بالدوران إلى الجانب الأيمن ثم بدأ بالمشي، أي تحرك بجسده إلى الأسفل ثم عاد بحركة عكسية إلى الأعلى ثم بدأ بالدوران، عندما يهبط جسد الرجل إلى الأسفل يستقر في وضعية القرفصاء لبرهة وهذا ما يجعل الجسم ساكناً بجميع نقاطه مميزاً وضعية القرفصاء كوضعية مرحلية عابرة ساكنة أثناء الحركة، بعد اعتدال الرجل يقوم بالدوران إلى اليمين مستخدماً جانبه الأيمن كمحور دوران وهذا ما يجعل هذا الجانب أكثر ثباتاً أثناء الحركة، بالنظر إلى صور MHI المحسوبة لهذا المشهد من أجل قيم مختلفة لـ τ (20,40,120,255)، فإننا نجد أن صور MHI لم توفق في تمثيل الحركة بسبب التغطية الذاتية التي تعاني منها، فهي لم تستطع تمثيل إلا حركة المشي الأخيرة، وهذا لأنها غطت على الحركات السابقة، أما صورة STTFI لذات المشهد من أجل $T=0.001$ ، فقد استطاعت تمثيل جوانب النشاط كلها فوضعية القرفصاء تظهر بسطوح كبير، كذلك الجانب الأيمن من الشخص يظهر كعمود مضيء وسط جسده معبراً عن محور الارتكاز أثناء الدوران، كما مثلت أيضاً حركة المشي الأخيرة.



مجموعة إطارات فيديو لشخص ينحني فيعتدل



MHI($\tau=20$)



MHI($\tau=40$)



MHI($\tau=120$)



MHI($\tau=255$)

في هذا النشاط تعاني صور MHI من التغطية الذاتية حيث صور الشخص الصاعد تغطي على صور الشخص المنحني القديمة



STTFI with KT($T=0.001$)

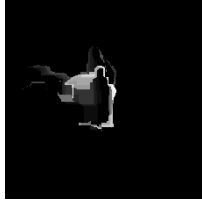
صور STTFI يلاحظ أن الانحناء والصعود بائن من الخط العامودي العريض المضيء ضمن جسم الشخص في الصورة، ووضع القرفصاء هو الأكثر إضاءة

الشكل 2—5 صور STTFI و MHI لنشاط الانحناء ثم الاعتدال

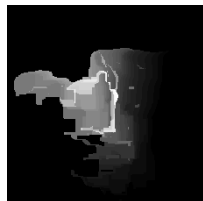
يظهر الشكل (3-5) مشهد لشخص يمشي ثم يجلس، في هذا النشاط تكون مشية الشخص على سرعات متفاوتة وبالتالي هناك وضعيات مرحلية عابرة تكون فيها هيئة الشخص بطيئة ويمكن أن نقول عليها أكثر استقراراً أو شبه ساكنة مقارنة بالوضعيات السريعة، كما أن الشخص عندما يصل إلى الأريكة يستدير مستخدماً الجانب الأيسر من جسمه مرتكزاً على الرجل اليسرى كمحور استدارة كما يظهر في المشهد. لم تستطيع صورة MHI مع القيم المختلفة لـ τ تمثيل ذلك إنما مثلت اتجاه الحركة فقط بالذهاب نحو الأريكة. بينما استطاعت صورة STTFI تمثيل ذلك بدقة، إذ يلاحظ فيها أن الشخص الماشي يمر بمواضع معينة تكون فيها هيئته أكثر إضاءة دالة على ببطء حركة المشي بينما يمر بأماكن تكون هيئته فيها أخف إضاءة دالة على أن المشية فيها سريعة، بينما يظهر واضحاً محور الارتكاز عند دوران الشخص لكي يجلس على الأريكة، حيث يظهر الجانب الأيسر من هيئته بإضاءة ساطعة.



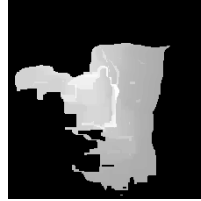
مجموعة إطارات فيديو لشخص يمشي فيجلس



MHI($\tau=20$)



MHI($\tau=40$)



MHI($\tau=120$)



MHI($\tau=255$)

يلاحظ أن صور MHI لم تميز مركز دوران الشخص عند استدارته لكي يجلس وهو الجانب الأيسر من جسم الشخص، إنما ميزت كيفية الحركة



STTFI with KT($T=0.001$)

صور STTFI ميزت مركز دوران الشخص في الجانب الأيسر من جسمه وذلك بنقاط أكثر إضاءة، كما ميزت مراحل تطور الحركة.

الشكل 3—5 صور MHI وصورة STTFI لشخص يمشي ثم يجلس

يبين الشكل (4-5) مشهد سقوط شخص نحو الخلف، حيث أثناء السقوط تكون قدمه اليمنى محور دوران حتى ارتطام ركبته بالأرض حيث يتوقف برهة (الإطار الثالث انطلاقاً من اليسار) ثم يتابع الدوران حول ركبته حتى يرتطم جذعه بالأرض عندها ترتفع القدم اليمنى عن الأرض، ويصبح الجذع محور الدوران الجديد، حيث يتوقف جسد الشخص لبرهة بعيد الارتطام وقبل متابعة الدوران حول الجذع، تنتهي عملية السقوط بالاستلقاء التام على الأرض، تظهر صور MHI حركات أطراف الشخص، أما اتجاه الحركة فيختفي تماماً نتيجة سلبية التغطية الذاتية التي تعاني منها صور MHI، كما أن وضع التوقف المرحلي عند ارتطام الركبة بالأرض مختفٍ تماماً. أما صورة STTFI فتظهر بوضوح وضع التوقف المرحلي هذا (الإطار الثالث انطلاقاً من اليسار)، كما تظهر بإضاءة ساطعة محاور الدوران المتعددة التي مر بها الشخص أثناء السقوط.



مجموعة إطارات فيديو لشخص يسقط إلى الخلف

MHI($\tau=20$)MHI($\tau=40$)MHI($\tau=120$)MHI($\tau=255$)

يلاحظ أن صور MHI لم تميز مركز دوران الشخص عند سقوطه على الأرض وملامسة جذعه لها

STTFI with KT($T=0.001$)

صور STTFI ميزت مركز دوران الشخص المؤلف من عدة محاور ارتكاز

الشكل 4—5 صور MHI و STTFI لشخص يسقط إلى الوراء

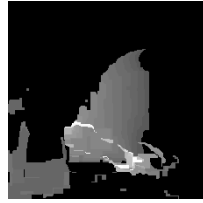
يبين الشكل (5-5) مشهد سقوط شخص إلى الأمام، إذ تستخدم الرجل اليسرى كمحور دوران، في البداية يبدأ السقوط بالارتكاز على القدم اليمنى، ثم ترتطم ركبة الشخص بالأرض فيتوقف برهة (الإطار الرابع انطلاقاً من اليسار)، بعدها تصبح الركبة محور دوران حتى الاستقرار على الأرض، تظهر صور MHI اتجاه الحركة وتتابعها حتى الاستقرار على الأرض، وذلك لأن اتجاه الحركة ليس موازياً لمحور الكاميرا في أغلب مراحل الحركة، وبالتالي لا تظهر سلبية التغطية الذاتية، لكن صور MHI لم تظهر محاور الدوران، ولا وضعية التوقف المرحلي الممثلة في الإطار الرابع، في حين أن صورة STTFI استطاعت إظهار محاور الحركة بإضاءة ساطعة، كما مثلت حالة التوقف المرحلي العابر كما في الإطار الرابع بدقة، إضافة إلى تمثيل اتجاه الحركة وأيضاً الحالة النهائية عند الاستلقاء مع التأذي، إذ ظهرت بإضاءة ساطعة دلالة على طول سكون الشخص على الأرض بعد الاستلقاء.



مجموعة إطارات فيديو لشخص يسقط إلى الأمام



MHI($\tau=20$)



MHI($\tau=40$)

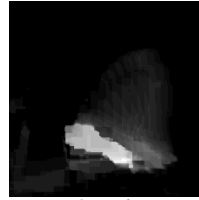


MHI($\tau=120$)



MHI($\tau=255$)

يلاحظ أن صور MHI لم تميز مركز دوران الشخص عند سقوطه على الأرض الذي هو الساق اليسرى وقدمها



STTFI with KT($T=0.001$)

صور STTFI ميزت مركز دوران الشخص المؤلف من الساق اليسرى وقدمها

الشكل 5-5 صورة STTFI وصور MHI لشخص يسقط إلى الأمام

5-3- منظومة كشف السقوط المقترحة باستخدام صور STTFI

سيتم استخدام صورة STTFI المقترحة الجديدة في منظومة لكشف السقوط يبينها الشكل (5-6)، إذ

تتألف المنظومة من أربعة مراحل: استحصال الأطر من الفيديو، وكشف هيئة الإنسان، واستخلاص صورة

STTFI، والشبكة العصبونية الالتفافية.

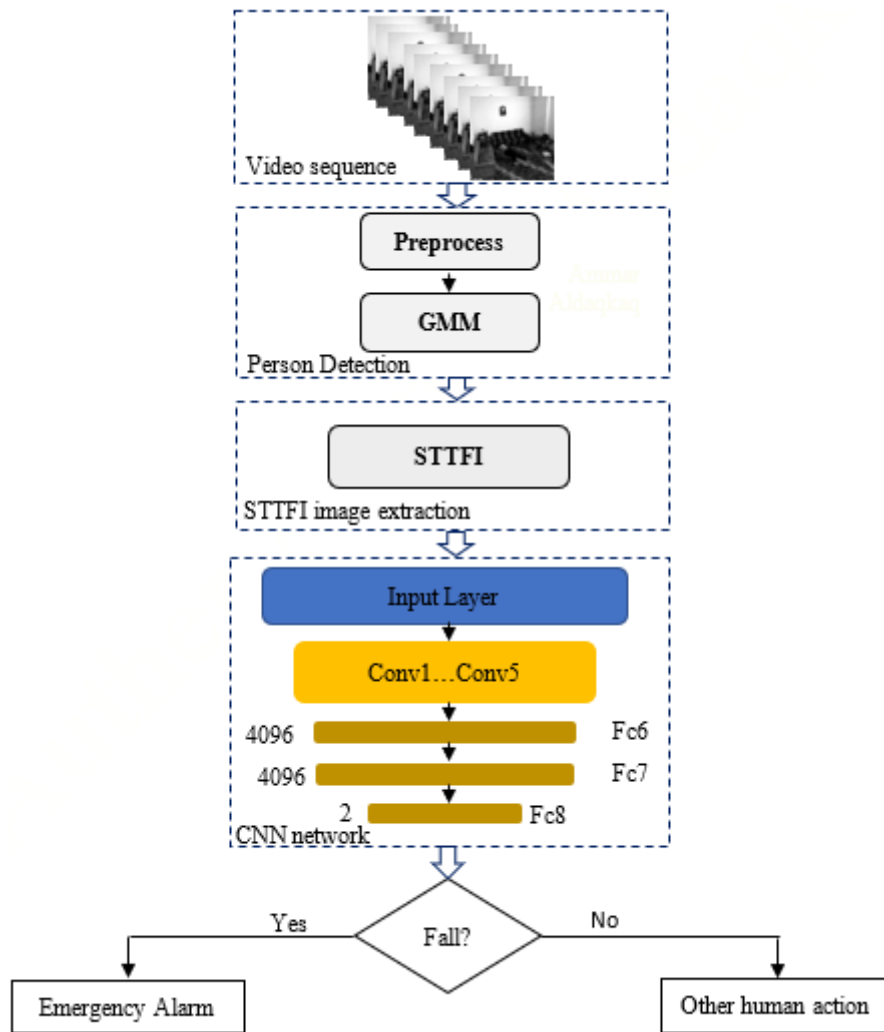
5-3-1- استحصال الأطر من الفيديو

يتم في هذه المرحلة استحصال الإطارات من الفيديو بزمان اعتيان ثابت مقداره 0.04s وهو يقابل معدل الإطار

25f/s، وبهذه الطريقة مهما كبر معدل الأطر الذي النقط فيه الفيديو، فإن الأطر المستحصلة تظهر وكأنها

سجلت بمعدل إطار 25f/s، يتم بعدها تجميع 120 إطارا لإرسالها إلى المرحلة اللاحقة، فقد وجد [32] نتيجة

متابعة العديد من الفيديوهاات لحالات السقوط بأن أطول فترة لحدوث السقوط هي أربعة ثوان (أي قد يتعثّر الشخص ثم يحاول التوازن ثم يسقط ولا يستطيع النهوض بسبب التأذي) بينما يتم السقوط الفعلي من الزلل إلى الاستلقاء في فترة تتراوح بين 300ms و 500ms، وعليه فإن 15 إطاراً كحد أقصى تكافئ فترة السقوط الفعلي، بينما السقوط بشكل عام قد يستغرق كحد أقصى زمن 120 إطاراً.



الشكل 5-6 منظومة كشف السقوط المقترحة باستخدام صورة STTFI.

5-3-2- كشف هيئة الإنسان Person detection

تتألف هذه المرحلة من قسمين: القسم الأول هو المعالجة الأولية، ويتم فيها تعديل حجم الأطر المستحصلة إلى قياس $227 \times 227 \times 3$ ، باستخدام الهرم الغوسي، كما يتم في هذه المرحلة تحويل الصور إلى

الرمادي، أما القسم الثاني فيتم فيه استخدام نموذج المزيج الغوسي (Gaussian Mixture Model (GMM لفصل هيئة الشخص عن خلفية الصورة، حيث يتم الحصول في نهاية هذه المرحلة على مجموعة صور لهيئة الشخص بدون الخلفية وبحجم مناسب لتكون دخلاً للمرحلة التالية.

3-3-5- استخلاص صورة STTFI

في هذه المرحلة يتم أخذ مجموعة الصور الناتجة من المرحلة السابقة، وتشكيل صورة STTFI الجديدة المقترحة التي شُرحت سابقاً، حيث تحتوي هذه الصورة على سمات مكانية وزمانية ليس فقط بما يتعلق بالحركة في المشهد بل حتى بما يتعلق بالنقاط الساكنة والساكنة مرحلياً في المشهد، وبالتالي فإن هذه الصورة تحتوي مواصفات تمييزية في وصف الفعل الحركي الذي يقوم به الشخص، وهذا ما سيظهر في النتائج اللاحقة.

5-3-4- الشبكة العصبونية الالتفافية CNN

في هذه المرحلة يتم ادخال صورة STTFI على الشبكة العصبونية الالتفافية لكي تقوم بمهمتين؛ الأولى استخلاص السمات المكانية والزمانية من الصورة، أما الثانية فهي القيام بتصنيف السمات المستخلصة إلى حدث سقوط أو عدم سقوط، فإن كان الحدث سقوط سيتم استدعاء المساعدة بإحدى وسائل الاتصال كرسائل SMS أو رسائل واتس أو ما شابه ذلك مع تحديد موقع المصاب بالسقوط، أما إن لم يكن الحدث سقوط تابعت المنظومة عملها في مراقبة المشهد.

اعتمد في هذه المنظومة على شبكة Alexnet التي تتمتع ببساطة التركيب، قلة العمق، قوة التمييز، وهي أول شبكة اقترحت في مجال التعلم العميق، كما ذكر في المقطع 2-1-4-1.

تم تعديل بنية شبكة Alexnet لتلائم الغرض من المنظومة وهو التعرف على حدث السقوط أو عدم السقوط، ولذلك فقد حُذفت طبقة FC8 التي تتعرف على 1000 صنف واستُبدلت بطبقة كاملة الاتصال FC8 جديدة بعصبونين للتعرف على أحد الحدثين. عند تعليم الشبكة سيتم فقط التعليم من الصفر لهذه الطبقة بينما

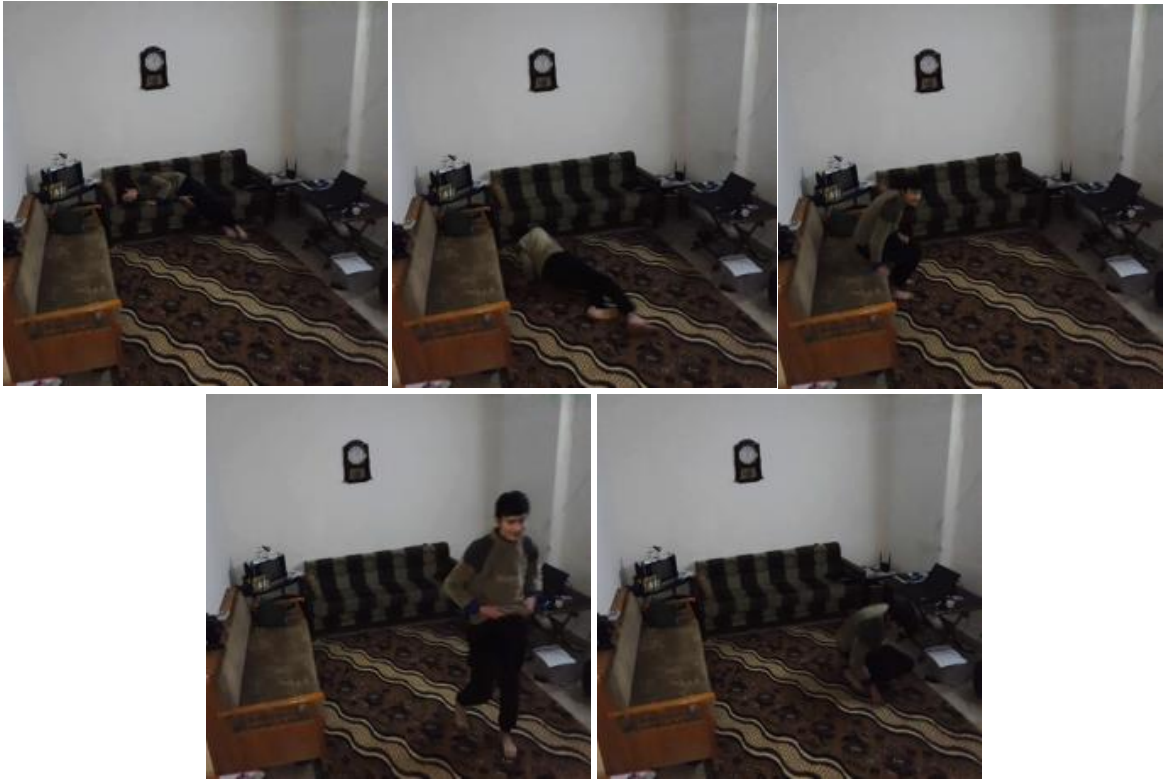
يتم تحديث قيم الأوزان في الطبقات الأخرى بما يتناسب مع مجموعات التدريب الجديدة وهذا جوهر نقل التعلم الذي من خلاله تحافظ الشبكة على ما تعلمته سابقاً وتتعلم بأمثلة قليلة وبفترة قصيرة تصانيف جديدة.

5-4- مراحل تحضير المعطيات والتدريب

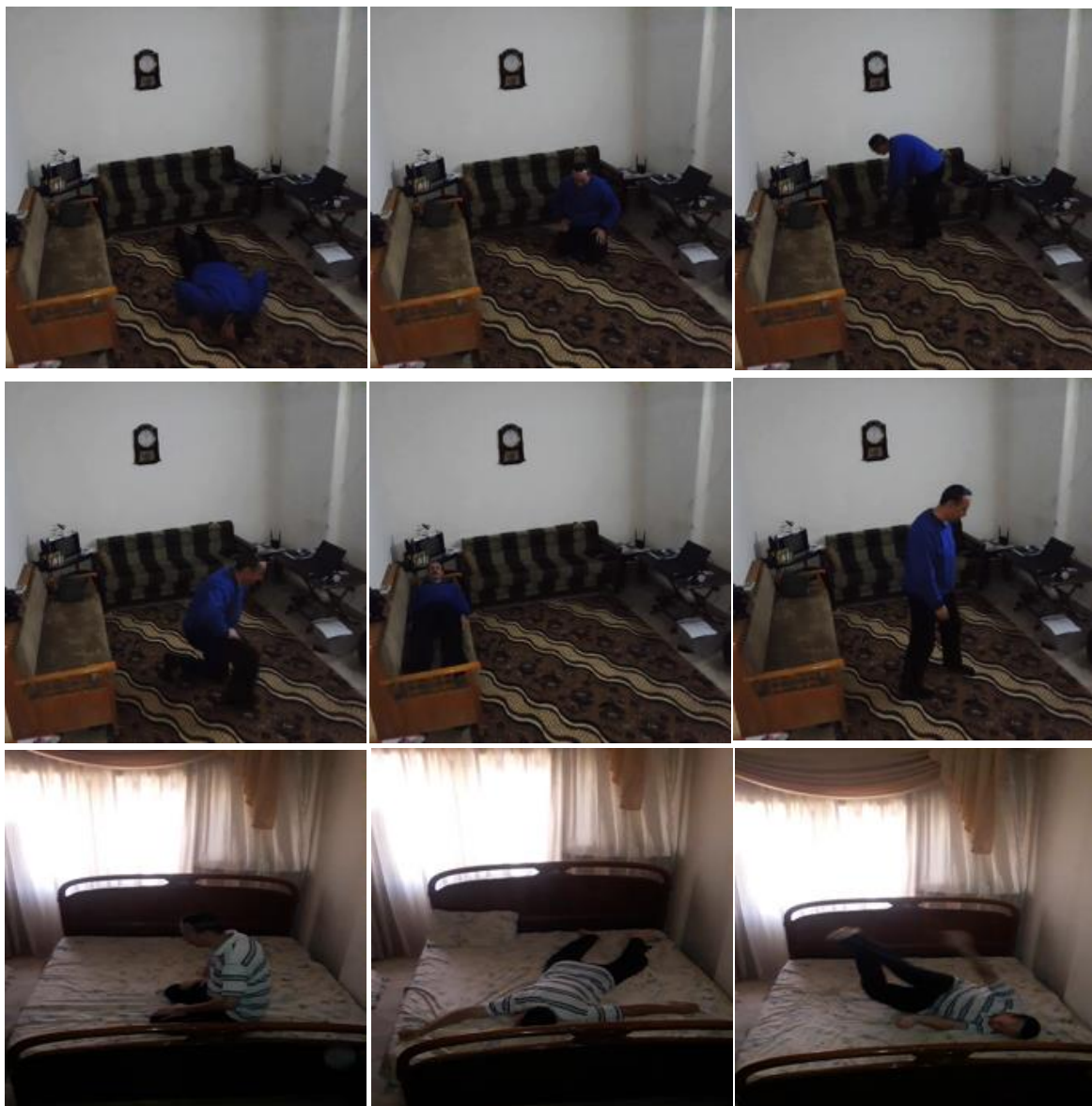
5-4-1- معطيات التدريب

لقد تم تدريب الشبكة على مجموعات صور تدريب هي:

1. مجموعة FDD ذكرت في الفقرة 4-3-1.
2. مجموعة URFD ذكرت في الفقرة 4-3-1.
3. مجموعة U. Montreal ذكرت في الفقرة 4-3-1.
4. مجموعة التدريب المصورة الجديدة FD22 الخاصة بالبحث، تتألف من 225 مشهداً - 47 مشهد سقوط والباقي نشاط حياتي يومي - مصورة بأبعاد 960x720 وبمعدل إطار 30 f/s، وذلك في كل من غرفة الجلوس وغرفة النوم، ومن خلال متطوعين، وتحتوي على 25485 صورة، منها 5297 صورة سقوط بمختلف الوضعيات، بما فيها صور جديدة للسقوط نتيجة فقدان الوعي، وكذلك السقوط من على الأريكة، ليست موجودة في المجموعات السابقة، يبين الشكل (5-7) والشكل (5-8) عدة صورة من مجموعة التدريب الجديدة FD22.



الشكل 5—7 صور من مجموعة التدريب FD22 لأحد المتطوعين.



الشكل 5—8 صور من مجموعة التدريب FD22 لأحد المتطوعين.

5-4-2- تحضير المعطيات

في هذه المرحلة تم تقسيم المشاهد إلى لقطات، ثم طبقت المرحلتين الثانية والثالثة من الخوارزمية ليتم الحصول في نهايتهما على صور STTFI، حيث حُفظت الصور التي تمثل أحداث السقوط في مجلد يحمل الاسم "fall"، بينما حُفظت الصور التي تمثل النشاط الحياتي اليومي في مجلد يحمل الاسم "not_fall".

5-4-3- مرحلة التدريب

في هذه المرحلة حُذفت الطبقة FC8 التي تميز بين 1000 صنف، وأضيفت طبقة FC8 مؤلفة من صنفين فقط وهما الصنفان اللذان ستميزهما شبكة Alexnet العصبونية الالتفافية، ثم تم تدريب الشبكة المعدلة بتقنية نقل التعلم، باستخدام الماتلاب 2020a، قُسمت مجموعة معطيات التدريب إلى 70% للتدريب و15% للاختبار و15% للصلاحيّة، أما بالنسبة لمعدل التعليم فقد جُعل بمقدار 0.001، أما بالنسبة لعدد دورات التدريب فقد اختيرت لتساوي 45 دورة، أما حجم الدفعة من الصور فهو 64.

تمت مرحلة التدريب على جزأين: الجزء الأول كان الغرض منه التأكد من أن الصورة المقترحة STTFI أفضل من صورة MHI في تمثيل النشاط بما فيه الوضعيات المرحلية العابرة ومحاور الارتكاز، فقد دُرِبَت Alexnet المعدلة على صور STTFI مستنتجة من مجموعة المعطيات FD22، وسُميت الشبكة المعدلة المدربة Alexnet_STTFI، دُرِبَت Alexnet المعدلة على صور MHI مستنتجة من ذات مجموعة التدريب وسُميت الشبكة المعدلة المدربة Alexnet_MHI، ثم قورنت نتائج الاختبار على الشبكتين لمعرفة هل صور STTFI المقترحة أفضل من صور MHI، في تمثيل نشاط السقوط.

أما الجزء الثاني فقد دُرِبَت فيه Alexnet المعدلة على صور STTFI المستنتجة من مجموعات التدريب الأربعة جميعها ثم قورنت النتائج مع الدراسات السابقة.

5-5- النتائج والمناقشة

تم تقييم أداء الشبكة من خلال معايير الأداء الرئيسية الثلاثة المذكورة في المقطع (1-7) وهي الدقة (Accuracy) والحساسية (Sensitivity) والنوعية (Specificity)، حيث تقيس الدقة هنا قدرة الطريقة المقترحة على التمييز بين أحداث السقوط وأنشطة الحياة اليومية، بينما تحدد الحساسية قدرة الطريقة المقترحة على تصنيف أحداث السقوط بشكل صحيح، أما النوعية فهي تحدد قدرة تلك الطريقة المقترحة على التمييز بين أنشطة الحياة اليومية بشكل صحيح.

يبين الجدول (1-5) مصفوفة الالتباس (confusion Matrix) المتعلقة بتعليم الشبكة الالتفافية المعدلة على صور MHI للجزء الأول من التدريب، كما يبين الجدول (2-5) مصفوفة الالتباس المتعلقة بتعليم الشبكة الالتفافية المعدلة على صور STTFI للطريقة المقترحة.

الجدول 1-5 مصفوفة الالتباس المتعلقة بشبكة Alexnet المدربة على صور MHI.

الوضع الحقيقي	عدد صور الاختبار	الوضع المُتنبأ به	
		سقوط	عدم سقوط
السقوط	10	9	1
عدم السقوط	26	2	24

الجدول 2-5 مصفوفة الالتباس المتعلقة بشبكة Alexnet المدربة على صور STTFI.

الوضع الحقيقي	عدد صور الاختبار	الوضع المُتنبأ به	
		سقوط	عدم سقوط
السقوط	10	10	0
عدم السقوط	26	1	25

ومن الجدولين السابقين يمكن حساب كل من المعايير الثلاثة السالفة الذكر، حيث يبين ذلك الجدول

(3-5).

الجدول 5-3 مقارنة بين الطريقة المقترحة وطريقة MHI

الطريقة	الحساسية	النوعية	الدقة
	Sensitivity	Specificity	Accuracy
MHI	90%	92.31%	91.66%
STTFI	100%	96.15%	97.22%

يلاحظ من الجدول (3-5) بأن دقة الشبكة باستخدام الطريقة المقترحة أعلى منها باستخدام طريقة MHI، حيث وصلت دقة الشبكة بالطريقة المقترحة إلى 97.22%، بينما باستخدام MHI وصلت دقة الشبكة ذاتها إلى 91.66%، وكذلك يلاحظ أن الشبكة العصبونية الالتفافية باستخدام الطريقة المقترحة STTFI استطاعت تمييز كل حركات السقوط المختبرة، بينما الشبكة العصبونية الالتفافية باستخدام صورة MHI وصلت الحساسية فيها إلى 90% فقط، وعليه فطريقة STTFI المقترحة الجديدة أفضل في تمثيل الأفعال والحركات البشرية من طريقة MHI، وهذا ما يثبت صحة طرح هذه الطريقة فهي تستطيع تمثيل الحركة التي يمر بها الجسم البشري أثناء قيامه بفعل معين، إضافة إلى تمثيل اللحظات الزمنية التي تكون فيها أجزاء من الجسم ساكنة مرحلياً أو كلياً، بينما MHI والطرق الأخرى المعدلة عنها تمثل بشكل جيد كمية الحركة واتجاهها ولا تهتم بالأجزاء الساكنة أو الساكنة مرحلياً، والتي تحمل سمات زمنية ومكانية هامة في تحديد نوع الحركة أو الفعل البشري.

أما فيما يتعلق بالجزء الثاني من التدريب فإن الجدول (4-5) يبين مصفوفة الالتباس للشبكة المعدلة المدربة على المجموعات التدريبية الأربعة جميعها.

الجدول 4-5 مصفوفة الالتباس لشبكة Alexnet المدربة على المجموعات الأربعة باستخدام صور STTFI.

الوضع الحقيقي	عدد صور الاختبار	الوضع المُتنبأ به	
		سقوط	عدم سقوط
السقوط	69	69	0
عدم السقوط	237	3	234

يُلاحظ من الجدول (4-5) أن الدقة النهائية التي تم الحصول عليها هي 99.02% والحساسية والنوعية

هما 100% و 98.73% على التوالي. يبين الجدول (5-5) مقارنة الطريقة المقترحة مع الطرق المعتمدة على

الشبكات الالتفافية وصور ملونة أو صور تدفق بصري.

الجدول 5-5 مقارنة بين الطريقة المقترحة والطرق التي تستخدم الشبكة العصبونية الالتفافية مع صور ملونة أو تدفق

بصري.

الطريقة	الشبكة المستخدمة	نوع الصور المستخدمة	مجموعة معطيات التدريب الصورية	الحساسية Sensitivity %	النوعية Specificity %	الدقة Accuracy %
Yu وآخرون (2017) [28]	شبكة CNN مقترحة	صور ثنائية اللون	مجموعة صورية خاصة به	98.98	99.03	99.01
Adhikari وآخرون (2017) [29]	شبكة CNN مقترحة	RGB-D	مجموعة صورية خاصة به	99.73	64.44	74
Anishchenko (2018) [30]	Alexnet	RGB	FDD	61	99	99
Núñez-Marcos وآخرون (2017) [31]	VGG-16	OF	URFD U.Montreal FDD	100 99 99	92 96 97	95 - 97
Espinosa وآخرون (2019) [33]	شبكة CNN مقترحة	2- OF	UP-Fall Detection	97.95	83.08	95.64
الطريقة الثانية المقترحة	Alexnet	STTFI	URFD, FDD, U.Montreal, FD22 خاصة بنا	100	98.73	99.02

يُلاحظ من الجدول (5-5) بأن الطريقة المقترحة تتفوق على جميع تلك الطرق السابقة، فرغم أن Yu وآخرين [28] وصل إلى دقة عالية 99.01%، الطريقة المقترحة أعطت دقة 99.02% أعلى بقليل من تلك الطريقة إلا أن الحساسية أي تمييز حدث السقوط بالنسبة للطريقة المقترحة هي 100%.

كما أن الطريقة المقترحة تفوقت من حيث الدقة والحساسية على طريقة Adhikari وآخرين [29] التي استخدم فيها صور RGB-D، يعود السبب في هذا إلى قلة العينات المستخدمة إضافة إلى محدودية السمات في صور RGB المستخدمة ونوعية شبكته الالتفافية، أما Anishchenko [30] فقد كانت الحساسية في طريقتها منخفضة جداً لأنها استخدمت مجموعة تدريب واحدة وصور RGB فقط. أما بالنسبة Núñez-Marcos [31] فإن الطريقة المقترحة تفوقت عليه من حيث الدقة والحساسية والنوعية رغم استخدامه شبكة التفافية أكثر عمقاً وتعقيداً. وقد تفوقت الطريقة المقترحة على طريقة Espinosa وآخرين [33] والتي استخدم فيها صور التدفق البصري التي تعبر فقط عن سرعة الحركة.

يبين الجدول (5-6) مقارنة بين الطريقة المقترحة والطرق المعتمدة على تعدد السيول multi-stream التي يقصد بها عدة شبكات التفافية تعمل على التفرع معاً وتأخذ صور مختلفة الصيغ، وهذا يعني شبكة ذات حجم كبير، إذ يلاحظ أيضاً من الجدول (5-6) ورغم أن أضخم شبكة عصبونية التفافية والمؤلفة من أربعة سيول كما هو ظاهر مع Khraief وآخرين [11] وصلت إلى حساسية 100% مع مجموعة المعطيات URFD، إلا أن الطريقة المقترحة حصلت على ذات الحساسية وبشبكة Alexnet الأقل عمقاً والأبسط تركيباً مع مجموعة معطيات متنوعة وأضخم عدداً، كما أن النوعية في الطريقة المقترحة أعلى منها في تلك الطريقة، فقد وصلت النوعية في الطريقة المقترحة إلى 98.73%، بينما في تلك الطريقة وصلت إلى 95% مع مجموعة المعطيات URFD، وهذا يدل أن صور STTFI تحمل صفات مكانية وزمانية أكثر مما تمثله من صفات الصور الأربعة المستخدمة في تلك الطريقة. يمكن النظر إلى بقية الطرق في الجدول (5-6) ليلاحظ بأن الطريقة المقترحة هي الأفضل بينهم، ويعود ذلك لقدرة صور STTFI على تمثيل الفعل الحركي والمراحل الساكنة منه زمانياً ومكانياً مهما كان اتجاه الحركة بالنسبة لمحور الرؤية.

الجدول 5-6 مقارنة بين الطريقة المقترحة والطرق التي تستخدم الشبكة العصبونية الالتفافية مع تعدد السيول.

الدقة Accuracy %	النوعية Specificity %	الحساسية Sensitivity %	مجموعة معطيات التدريب الصورية	نوع الصور المستخدمة	الشبكة المستخدمة	الطريقة
-	92.5	100	URFD	OF BMI	2-streams VGG-16	Khraief وآخرون [35] (2019)
- -	98 99.3	98.8 100	Dataset-1 خاصة به Dataset-2 خاصة به	SI DI MHI	3-streams Resnet18	Kong وآخرون [36] (2019)
98.86 -	- 95	- 100	FDD URFD	OF BMI Disp.OF RGB-D	4-streams VGG-16	Khraief وآخرون [11] (2020)
99.02	98.73	100	URFD, FDD, U. Montreal, FD22 خاصة بنا	STTFI	Alexnet	الطريقة الثانية المقترحة

كما يبين الجدول (5-7) أيضاً مقارنة بين الطريقة المقترحة والطرق التي تعتمد على شبكة عميقة مؤلفة من مرحلتين متتابعين الأولى التفافية والثانية LSTM، إذ يلاحظ من الجدول بأن الطريقة المقترحة أفضل من كل الطرق الموجودة في الجدول إن كان من ناحية الدقة أو من ناحية الحساسية، فعند مقارنة الطريقة المقترحة مع طريقة Sangwan وآخرين [32] نجد بأن الطريقة المقترحة أفضل من ناحية الدقة، رغم استخدامه لصور سقفية بحيث تخلص من سيئة اختفاء أجزاء من الجسم وراء الأثاث، واستخدامه شبكة عميقة inceptionV3 تلتها شبكة LSTM، ولكن شعاع السمات لديه الذي يربط بين الشبكة الالتفافية وشبكة LSTM شعاع عمومي لا يمثل السمات بشكل دقيق كما نوقش في الفصل السابق، ومجموعة صور كانت من نوع RGB وقليلة العدد وتعتمد على شخص واحد.

الطريقة المقترحة أفضل من طريقة Lie وآخرين [34] من ناحية الدقة أيضاً، فالطريقة المقترحة باستخدام صور STTFI التي تحمل خصائص زمنية تتابعية ومكانية تعبر عن النشاط زمنياً ومكانياً، بينما الطرق الأخرى

تستخدم شبكة LSTM لكشف التتابع الزمني الرابط بين الأطر المتعاقبة، وهي أعمق، وبالتالي تستهلك موارد الحاسوب بشكل أكبر، كما أنها تعبر عن النشاط ككل وليس أجزاء النشاط المكونة له، كما أن شعاع السمات المستخدم في ذلك النوع من الشبكات عمومي لا يمثل النشاط بشكل دقيق. إذن باستخدام صور STTFI وبشبكة قليلة العمق تم الحصول على دقة أفضل من استخدام شبكات عميقة على مرحلتين.

كما يلاحظ أيضاً أن دقة الطريقة المقترحة أكبر من دقة شبكة Salimi [109] التي بلغت 98.9%، وأيضاً أكبر من دقة شبكة Li [110] التي بلغت 97.1%.

الجدول 5-7 مقارنة بين الطريقة المقترحة والطرق التي تستخدم شبكة عصبونية عميقة مؤلفة من مرحلتين متتابعتين.

الطريقة	الشبكة المستخدمة	نوع الصور المستخدمة	مجموعة معطيات التدريب الصورية	الحساسية Sensitivity %	النوعية Specificity %	الدقة Accuracy %
Sangwan وآخرون (2020) [32]	inceptionV3 and LSTM	RGB from ceiling	360°dataset خاصة به	–	–	91.62
Lie وآخرون (2019) [34]	CNN خاصة به مع LSTM	MHI	مجموعة خاصة به	–	–	97.66
Salimi وآخرون (2022) [109]	TD-CNN-LSTM	RGB	URFD	–	–	98.90
Li وآخرون (2022) [110]	DCNN and LSTM	Depth	URFD	–	–	97.1
الطريقة الثانية المقترحة	Alexnet	STTFI	URFD, FDD, U.Montreal, FD22 خاصة بنا	100	98.73	99.02

5-6- الخلاصة

في هذه الفصل طُرحت طريقة جديدة لكشف سقوط الإنسان بواسطة الكاميرا، تتلخص الطريقة في استخلاص صورة حركة جديدة سُميت STTFI تُختصر فيها مجموعة من إطارات الفيديو المتلاحقة، حيث تتضمن تلك الصورة معلومات زمانية ومكانية عن حركة الجسم البشري، وتمثل كيف حدثت الحركة ومتى حدثت،

إضافة إلى قدرتها على تمثيل الأجزاء الساكنة التي تمثل محاور ارتكاز أو الوضعيات الساكنة العابرة، وهذا لم يكن موجوداً في الطرق السابقة مثل MHI، فصورة STTFI تمثل كل الوضعيات المرحلية التي يمر بها الجسم البشري أثناء قيامه بحركة معينة إرادية أو لا إرادية، كما تستطيع تمثيل الحركة حتى وإن كان أحد محاور الحركة يوازي محور الرؤية؛ أي تخلصت من سلبية التغطية الذاتية في صور MHI. ثم تم استخدام تقنية التعلم العميق بواسطة الشبكات العصبونية الالتفافية، حيث اختيرت شبكة Alexnet لبساطتها وقلة عمقها ودقتها العالية في التصنيف، تم تعديلها لتستطيع التمييز بين صنفين هما السقوط وعدم السقوط، وتم تصوير مجموعة معطيات تدريب خاصة FD22 تحتوي وضعيات جديدة مثل حالة الإغماء، تم بعدها تدريب الشبكة بواسطة نقل التعلم على صور STTFI المستنتجة من مجموعات المعطيات الأربعة جميعها (URFD, FDD, U. Montreal,)، أعطت الطريقة المقترحة نتائج أفضل من الأبحاث السابقة في هذا المجال، حيث أعطت دقة 99.02% وحساسية 100% ونوعية 98.73%.

الخاتمة

في هذا البحث، تمت دراسة ومناقشة نشاط من أهم النشاطات الحياتية التي تحصل مع الإنسان ألا هو السقوط، وذلك لما فيه من أثر بالغ على صحته خاصة كبار السن وصغار السن، وقد تم سرد العديد والعديد من الدراسات التي تحدثت عن ذلك في العقد الأخير، وتمت مناقشة تلك الدراسات والطرق، والتي انقسمت إلى طرق تستخدم الحساسات المحيطية وأخرى تستخدم الحساسات الملاصقة للجسم وأخرى تستخدم الإبصار الحاسوبي الذي جعله ظهور الكاميرات وانتشارها بكثرة وتطور طرق الإبصار الحاسوبي في الصدارة خلال العقد الأخير، وقد انقسمت طرق الإبصار الحاسوبي إلى طرق تقليدية تعتمد استخراج سمات الصور بشرياً من خلال الطرق الهندسية والإحصاء الرياضياتي ثم استخدام العتبات أو تعليم الآلة لتصنيف الحدث، وطرق التعلم العميق التي تستخلص الخصائص تلقائياً، إذ كان لها الأثر الفعال في الدراسات التي حصلت في السنوات الأخيرة.

اعتمدت طرق التعلم العميق على الشبكات الالتفافية في كشف حدث السقوط، أو على شبكات هجينة تدمج الشبكات الالتفافية مع شبكات LSTM، حيث تقوم الأولى بتمييز السمات المكانية في الصور بينما الثانية تقوم بتمييز الترابط الزمني بين تلك السمات المتعاقبة، ولكن الربط بينهما كان من خلال الطبقة ما قبل الأخيرة في الشبكات الالتفافية، حيث تحمل تلك الطبقة سمات أكثر تعميماً لنشاطات محددة منها الاستلقاء، وعلى هذه الحال فإن الشبكة الهجينة تحاول الربط الزمني بين النشاطات المختلفة باستخدام سمات معممة، وهذا ما يعطي اشكاليتين؛ الأولى هي التشابه بين الوضعيات المرحلية الجزئية للنشاطات المختلفة مثل الانحناء والسقوط وهذا ما يضعف دقة الكشف حتى بعد التعليم على مجموعة كبيرة من الصور، أما الثانية فهي استخدام سمات عمومية لمعرفة النشاط فسقوط عن سقوط يفرق، لحل هاتين الإشكاليتين، تم القيام بتمييز فقط حدث السقوط وقُسم إلى ثلاثة مراحل جزئية، وجُعِلَت الشبكة الالتفافية تكشف تلك المراحل الثلاثة، ولتمييز تلك المراحل اختير شعاع السمات من طبقتين الأولى ذات سمات تفصيلية والثانية ذات سمات عمومية، عندها تكون مهمة شبكة LSTM

هي كشف الربط الزمني بين المراحل الثلاثة للسقوط وبناء على شعاع السمات المختار، وهذا ما أعطى نتائج جيدة وتمييزية عالية لحدث السقوط. وهذا كان صلب المنظومة الأولى التي تم اقتراحها.

أما الطرق التي كانت تعتمد فقط على الشبكات الالتفافية، فقد كان الاعتماد في كشف السمات المكانية يتم إما على الأطر الملونة المستخرجة من الفيديو، أو على الأطر الملونة والأطر ذات العمق من كاميرات النقاط العمق، أو على صور لهيئة الشخص المستخرجة من الفيديو، وقد انتهجت هذه الأخيرة إما تمييز الحركة فقط أو تمييز الوضعيات الساكنة فقط، وكان أشهر تلك الطرق هي صورة MHI والطرق المطورة عليها، وهذه الصورة تصف الحركة التي جرت خلال المشهد من حيث كيف حدثت وتجاهلت متى حدثت، كما أنها جميعها تهتم بما طرأ أخيراً على الحركة فمثلاً في بكسلات أغلب الطرق تكون البكسلات التي جرت عليها الحركة مؤخراً أكثر سطوعاً من تلك التي مرت عليها الحركة سابقاً، بينما تتعرض البكسلات ذات الحركة القديمة إما للاضمحلال أو للتغطية من البكسلات الجديدة، إضافة إلى عدم ربط الزمن بشكل مناسب معها؛ ولحل تلك الإشكالية تم طرح طريقة جديدة تدمج السمات المكانية والزمانية المرتبطة بها، مع تمييز الوضعيات المرحلية الساكنة التي تمر بها أجزاء من الجسم أثناء السقوط مع اللحظات الزمنية التي تكون فيها تلك الوضعيات ساكنة، وقد سُميت الطريقة الجديدة صورة أمامية مكسدة بعلامات زمنية (STTFI) Stacked Time-Tagged Foreground Image، تم استخدام هذه الطريقة ضمن منظومة تستخلص هيئة الشخص ثم تستخدم عدة صور متعاقبة لها لتكوين صورة STTFI التي تستخدم كدخل لشبكة التفافية تقوم بتمييز حدث السقوط من خلال تحليل السمات الزمنية والمكانية التي تحويها صورة STTFI لحدث السقوط بوضعياته المرحلية المتحركة والساكنة، وقد أعطت تلك الطريقة تمييزاً تاماً لحدث السقوط المؤذي الذي يبقى الشخص فيه ممدداً على الأرض بعد السقوط لعدم قدرته على القيام إما لكسر أو لغياب عن الوعي.

6-1- الاستنتاجات

1. النظر إلى السقوط كوضعيات مرحلية جزئية متعددة يعطي نتائج ودقة أفضل.

2. التركيز على الوضعيات المرحلية الجزئية المشكلة للنشاط المراد كشفه فقط يعطي نتائج ودقة أفضل.
3. الربط بين الشبكات الهجينة CNN-LSTM بشعاع سمات عمومي تفصيلي يعطي نتائج أفضل.
4. صور MHI كانت قاصرة عن التعبير عن الوضعيات المرحلية الساكنة ومحاور الارتكاز.
5. صور MHI كانت قاصرة عن تمثيل الحركة عندما تكون أحد محاورها موازية لمحور الرؤية.
6. صور MHI تعاني من التغطية الذاتية.
7. صور MHI غير مناسبة للنشاطات المركبة الطويلة كالمشي ثم الجلوس... إلخ.
8. أثبتت صور STTFI قدرتها ودقتها في تمثيل الوضعيات المرحلية العابرة الساكنة، ومحاور الارتكاز الساكنة أثناء قيام الشخص بنشاط السقوط.
9. أثبتت صور STTFI قدرتها على تمثيل الحركة حتى وإن كان أحد محاورها موازياً لمحور الرؤية.
10. صور STTFI مناسبة لتمثيل النشاطات المركبة الطويلة كالمشي ثم الجلوس... إلخ.
11. يمكن الاستغناء عن الشبكات الهجينة CNN-LSTM واستبدالها بشبكة عميقة وحيدة في حال استخدام صور STTFI.
12. يمكن الاستغناء عن طرق تعدد السيول multi-stream واستبدالها بشبكة عميقة وحيدة في حال استخدام صور STTFI.
13. صور STTFI أقل اختراقاً لخصوصية المستخدم.

6-2- الأفاق المستقبلية

1. البحث في طرق أخرى لتطوير تقنية الربط بين CNN و LSTM، بحيث تقدم نتائج أفضل.
2. استخدام صور STTFI مع شبكات التفاضلية أخرى غير Alexnet لكشف حدث السقوط.

3. استخدام الطريقتين المقترحتين في التعرف أو تصنيف نشاطات أخرى غير السقوط، والتحقق من مدى دقتهما.
 4. تجربة الطريقتين على مجموعات صور جديدة، لم تكونا مجربتين عليها.
 5. البحث في أفاق استخدام الطريقتين في مجالات أخرى مثل التحكم في الروبوتات أو التشخيص الطبي للأمراض من خلال تحليل الحركات.
 6. وضع المنظومتين اللتين تعتمدان على الطريقتين المقترحتين في أماكن يوجد فيها مسنون كدور المسنين أو المشافي للتأكد من فعالية الطريقتين.
- في نهاية هذا البحث يوصى باتباع أحد النهجين عند اكتشاف نشاط معين قد يقوم به الإنسان خاصة تلك الأنشطة غير الطبيعية التي تحصل للإنسان مثل السقوط كحالة دراسية في هذا البحث.

المراجع

- [1] P. V. K. Borges, "Pedestrian Detection Based on Blob Motion Statistics," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 224 - 235, 2013.
- [2] A. ULLAH, J. AHMAD, K. MUHAMMAD, M. SAJJAD and S. W. BAIK1, "Action Recognition in Video Sequences using Deep Bi-Directional LSTM With CNN Features," *IEEEAccess*, pp. 1155 - 1166, 2017.
- [3] P. Bharti, D. De, S. Chellappan and S. K. Das, "HuMAN: Complex Activity Recognition with Multi-modal Multi-positional Body Sensing," *IEEE Transactions on Mobile Computing*, pp. 857 - 870, 1 April 2019.
- [4] A. F. Bobick and J. W. Davis, "The Recognition of Human Movement Using Temporal Templates," *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, pp. 257-267, Mar 2001.
- [5] D. Kim, H. Kim, Y. Mok and J. Paik, "Real-Time Surveillance System for Analyzing Abnormal Behavior of Pedestrians," *Applied Sciences MDPI*, pp. 2-16, 27 June 2021.
- [6] O. A. Elsayed, N. A. M. Marzouky, E. Atef and M. . A.-M. Salem, "Abnormal Action detection in video surveillance," in *IEEE Ninth International Conference on Intelligent Computing and Information Systems (ICICIS)*, 2019.
- [7] L. Alhimale, H. Zedan and A. Al-Bayatti, "The implementation of an intelligent and video-based fall detection system using a neural network," *Applied Soft Computing*, p. 59–69, 18 January 2014.
- [8] D. Cahoolessur and B. Rajkumarsingh, "Fall Detection System using XGBoost and IoT," *R & D Journal of the South African Institution of Mechanical Engineering*, pp. 8-18, 12 June 2020.

- [9] H. Li, C. Li and Y. Ding, "Fall detection based on fused saliency maps," *Multimedia Tools and Applications*, p. 1883–1900, 25 August 2021.
- [10] World Health Organization (WHO), "WHO," 26 April 2021. [Online]. Available: <https://www.who.int/ar/news-room/fact-sheets/detail/falls>. [Accessed 10 2 2022].
- [11] C. Khraief, F. Benzarti and H. Amiri, "Elderly fall detection based on multi-stream deep convolutional networks," *Multimedia Tools and Applications*, p. 19537–19560, 25 3 2020.
- [12] K. Chaccour, R. Darazi, A. H. el Hassans and E. Andres, "Smart Carpet using differential piezoresistive pressure sensors for elderly fall detection," in *IEEE 11th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2015.
- [13] M. Alwan, P. J. Rajendran, S. Kell, D. C. Mack, S. Dalal, M. Wolfe and R. Felder, "A Smart and Passive Floor-Vibration Based Fall Detector for Elderly," in *2nd International Conference on Information & Communication Technologies*, 2006.
- [14] Z. Z. Islam, S. M. Tazwar, M. Z. Islam, S. Serikawa and M. A. R. Ahad, "Automatic Fall Detection System of Unsupervised Elderly People Using Smartphone," in *5th IIAE International Conference on Intelligent Systems and Image Processing*, Japan, 2017.
- [15] S. RajuDhole, A. Kashyap, A. N. Dangwal and R. Mohan, "A novel helmet design and implementation for drowsiness and fall detection of workers on-site using EEG and Random-Forest Classifier," *Procedia Computer Science*, pp. 947-952, 21 May 2019.
- [16] D. Sarabia-Jácome, R. Usach, C. E. Palau and M. Esteve, "Highly-efficient fog-based deep learning AAL fall detection system," *Internet Of Things*, pp. 1-19, 26 sep 2020.
- [17] N. Thakur and C. Y. Han, "A Study of Fall Detection in Assisted Living: Identifying and Improving the Optimal Machine Learning Method," *Sensor and Actuator Network*, pp. 1-27, 24 June 2021.

- [18] A. Tahir, J. Ahmad, G. Morison, H. Larijani, R. . M. Gibson and D. A. Skelton, "HRNN4F:HYBRID DEEP RANDOM NEURAL NETWORK FOR MULTI-CHANNEL FALL ACTIVITY DETECTION," *Probability in the Engineering and Informational Sciences*, pp. 37-70, january 2021.
- [19] S. Albawendi, A. Lotfi, H. Powell and K. Appiah, "Video Based Fall Detection using Features of Motion, Shape and Histogram," in *the 11th Pervasive Technologies Related to Assistive Environments Conference*, 2018.
- [20] K. Sehairi, F. Chouireb and J. Meunier, "Elderly fall detection system based on multiple shape features and motion analysis," in *International Conference on Intelligent Systems and Computer Vision (ISCV)*, Fez, Morocco, 2018.
- [21] M. Ko, S. Kim, M. Kim and K. Kim, "A Novel Approach for Outdoor Fall Detection Using Multidimensional Features from A Single Camera," *applied sceinces*, pp. 1-22, 15 june 2018.
- [22] J. Thummala and S. Pumrin, "Fall Detection using Motion History Image and Shape Deformation," in *8th International Electrical Engineering Congress (iEECON)*, Chiang Mai, Thailand, 2020.
- [23] N. Worrakulpanit and P. Samanpiboon, "Human Fall Detection Using Standard Deviation of C-Motion Method," *Journal of Automation and Control Engineering*, pp. 388-391, December 2014.
- [24] K. Lakshmi and T. Devendran, "Human Fall Detection Using Motion History Image and SVM," in *International Conference on Advanced Informatics for Computing Research*, 2020.
- [25] C. Rougier, J. Meunier, A. St-Arnaud and J. Rousseau, "Fall Detection from Human Shape and Motion History using Video Surveillance," in *21st International Conference on Advanced Information Networking and Applications Workshops (AINAW'07)*, 2007.
- [26] M. B. Prasetyo, E. S. Wahyuni and H. Setiawan, "Comparison of Motion History Image and Approximated Ellipse Method in Human Fall Detection System," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, pp. 199-208, April 2019.

- [27] R. Dubey, B. ni and P. Moulin, "A Depth Camera Based Fall Recognition System for the Elderly," in *9th international conference on Image Analysis and Recognition - Volume Part II*, 2012.
- [28] M. Yu, L. Gong and S. Kollias, "Computer vision based fall detection by a convolutional neural network," in *19th ACM International Conference on Multimodal Interaction(ICMI)*, 2017.
- [29] K. Adhikari, H. Bouchachia and H. Nait-Charif, "Activity recognition for indoor fall detection using convolutional neural network," in *Fifteenth IAPR International Conference on Machine Vision Applications (MVA)*, Nagoya, Japan, 2017.
- [30] L. Anishchenko, "Machine learning in video surveillance for fall detection," in *2018 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBREIT)*, Yekaterinburg, Russia, 2018.
- [31] A. N.-. Marcos, G. Azkune and I. A.-. Carreras, "Vision-Based Fall Detection with Convolutional Neural Networks," *Wireless Communications and Mobile Computing*, pp. 1-16, 6 Dec 2017.
- [32] D. Sangwan, R. Manekar, S. Saurav, S. Maiti, S. Singh, S. Chaudhury, N. Kumar, R. Kumar and K. Chaudhary, "Activity Recognition for Indoor Fall Detection in 360-Degree Videos Using Deep Learning Techniques," in *3rd International Conference on Computer Vision and Image Processing*, New delhi, India, 2020.
- [33] R. Espinosa, H. Ponce, S. Gutiérrez, . L. M.-. Villaseñor, J. Brieva and E. M.-. Albor, "A vision-based approach for fall detection using multiple cameras and convolutional neural networks: A case study using the UP-Fall detection dataset," *Computers in Biology and Medicine*, 27 November 2019.
- [34] W.-N. Lie, F.-Y. Hsu and Y. Hsu, "Fall-down event detection for elderly based on motion history images and deep learning," in *International Workshop on Advanced Image Technology (IWAIT) 2019*, Singapore, 2019.
- [35] C. Khraief, F. Benzarti and H. Amiri, "Convolutional neural network based on dynamic motion and shape variationsfor elderly fall detection,"

- International Journal of Machine Learning and Computing*, pp. 814-820, 19 Dec 2019.
- [36] Y. Kong, J. Huang, S. Huang, Z. Wei and S. Wang, "Learning spatiotemporal representations for human fall detection in surveillance video," *Journal of Visual Communication and Image Representation*, pp. 215-230, 22 January 2019.
- [37] H. Ramirez, S. A. Velastin, I. Meza, E. Fabregas, D. Makris and G. Farias, "Fall Detection and Activity Recognition Using Human Skeleton Features," *IEEE Access*, pp. 33532-33542, 23 February 2021.
- [38] motionmetrics, "How Artificial Intelligence Revolutionized Computer Vision: A Brief History," 16 5 2019. [Online]. Available: <https://www.motionmetrics.com/how-artificial-intelligence-revolutionized-computer-vision-a-brief-history/#:~:text=Computer vision began in earnest,that could transform the world..> [Accessed 29 8 2022].
- [39] A. Krizhevsky, I. Sutskever and G. . E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in neural information processing systems* 25, 2012.
- [40] S. Haykin, *Neural Networks and Learning Machines*, Pearson, 2009.
- [41] A. D. Dongare, R. R. Kharde and A. D. Kachare, "Introduction to Artificial Neural Network," *International Journal of Engineering and Innovative Technology (IJEIT)*, pp. 1-6, 7 2012.
- [42] IBM, "Supervised Learning," 19 8 2020. [Online]. Available: <https://www.ibm.com/cloud/learn/supervised-learning#:~:text=Supervised learning%2C also known as,data or predict outcomes accurately..> [Accessed 3 9 2022].
- [43] D. . E. Rumelhart, G. . E. Hinton and J. L. McClelland, "A general framework for parallel distributed processing," in *Parallel distributed processing: Explorations in the microstructure of cognition*, 1986, pp. 45-76.
- [44] G. . F. Luger, in *Artificial Intelligence*, Mexico, Pearson, 2009.

- [45] M. Torma, "Kohonen self-organizing feature map and its use in clustering," *ISPRS Commission III Symposium: Spatial Information from Digital Photogrammetry and Computer Vision*, pp. 830-835, 1994.
- [46] D. . H. HUBEL and T. . N. WIESEL, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *The Journal of physiology*, pp. 106-154, 1962.
- [47] K. FUKUSHIMA, "Neocognitron: A hierarchical neural network capable of visual pattern recognition," *Neural networks*, pp. 119-130, 1988.
- [48] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, pp. 2278-2324, 9 1998.
- [49] stanford University, "Convolutional Neural Networks For Visual Recognition," 2022. [Online]. Available: <https://cs231n.github.io/convolutional-networks/>. [Accessed 17 9 2022].
- [50] stanford university, "Convolutional Neural Networks for Visual Recognition," 2022. [Online]. Available: <https://cs231n.github.io/convolutional-networks-1/>. [Accessed 17 9 2022].
- [51] S. Indolia, A. K. Goswami, S. . P. Mishra and P. Asopa, "Conceptual understanding of convolutional neural network-a deep learning approach," *Procedia computer science*, pp. 679-688, 2018.
- [52] Knowledge Transfer, "Explain Pooling layers: Max Pooling, Average Pooling, Global Average Pooling, and Global Maxpooling,," 29 8 2021. [Online]. Available: <https://androidkt.com/explain-pooling-layers-max-pooling-average-pooling-global-average-pooling-and-global-max-pooling/>. [Accessed 17 9 2022].
- [53] M. . M. Leonardo, T. J. Carvalho, E. Rezende, R. Zucchi and F. A. Faria, "Deep feature-based classifiers for fruit fly identification (Diptera: Tephritidae)," in *31st SIBGRAPI conference on graphics, patterns and images (SIBGRAPI)*, IEEE, 2018.

- [54] M. Hussain, J. J. Bird and D. . R. Faria, "A study on cnn transfer learning for image classification," in *In UK Workshop on computational Intelligence*, Cham, 2018.
- [55] و. الصافي، "أهم البنى المعمارية لتقنيات التعلم العميق"، 2022. [متصل]. Available: <http://www.scs.org.sy/?q=ar/node/893>. [تاريخ الوصول 18 9 2022].
- [56] S. Hochreiter and J. Schmidhuber, "Long short-term memory.," *Neural computation*, 9(8), pp. 1735-1780, 1997.
- [57] A. Amidi and S. Amidi, "CS 230 - Recurrent Neural Networks Cheatsheet," 2022. [Online]. Available: <https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-recurrent-neural-networks#cheatsheet>. [Accessed 19 9 2022].
- [58] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," *Advances in neural information processing systems*, 27, 2014.
- [59] Z. Wu, X. Wang, Y.-G. Jiang, H. Ye and X. Xue, "Modeling spatial-temporal clues in a hybrid deep learning framework for video classification," in *In Proceedings of the 23rd ACM international conference on Multimedia*, 2015.
- [60] J. Brownlee, "CNN Long Short-Term Memory Networks," 14 8 2019. [Online]. Available: <https://machinelearningmastery.com/cnn-long-short-term-memory-networks/>. [Accessed 20 9 2022].
- [61] J. Donahue, L. A. Hendricks, M. Rohrbach, S. Venugopalan, S. Guadarrama, K. Saenko and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description," in *In Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015.
- [62] X. Shi, Z. Chen, H. Wang and D.-Y. Yeung, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," *Advances in neural information processing systems*, 28, 2015.
- [63] T.-Y. Kim and s.-B. Cho, "Particle swarm optimization-based CNN-LSTM networks for forecasting energy consumption," in *In 2019 IEEE congress on evolutionary computation (CEC)*, 2019.

- [64] H. XIE, L. ZHANG and C. P. LIM, "Evolving CNN-LSTM models for time series prediction using enhanced grey wolf optimizer," *IEEE Access*, 8, pp. 161519-161541, 2020.
- [65] S. MONTAHA, S. AZAM, A. K. M. . R. H. RAFID, M. Z. HASAN, A. KARIM and A. ISLAM, "TimeDistributed-CNN-LSTM: A Hybrid Approach combining CNN and LSTM to classify brain tumor on 3D MRI scans performing ablation study," *IEEE Access* 10, pp. 60039-60059, 29 5 2022.
- [66] W. Boulila, H. Ghandorh, M. A. Khan, F. Ahmed and J. Ahmad, "A novel CNN-LSTM-based approach to predict urban expansion," *Ecological Informatics*, 64, 2021.
- [67] A. F. Gad, "Evaluating Deep Learning Models: The Confusion Matrix, Accuracy, Precision, and Recall," 2020. [Online]. Available: <https://blog.paperspace.com/deep-learning-metrics-precision-recall-accuracy/>. [Accessed 21 9 2022].
- [68] W. T. Freeman and M. Roth, "Orientation histograms for hand gesture recognition," in *In International workshop on automatic face and gesture recognition*, 12, 1995.
- [69] M. A. R. Ahad, J. . K. Tan, H. Kim and S. Ishikawa, "Motion history image: its variants and applications.," *Machine Vision and Applications*, 23(2), pp. 255-281, 2012.
- [70] D. Fortun, P. Bouthemy and C. Kervrann, "Optical flow modeling and computation: A survey," *Computer Vision and Image Understanding*, 134, pp. 1-21, 2015.
- [71] H. J. Lee and J. H. Chung, "Hand gesture recognition using orientation histogram," in *In Proceedings of IEEE. IEEE Region 10 Conference. TENCON 99. Multimedia Technology for Asia-Pacific Information Infrastructure'*, 1999.
- [72] J. W. Davis and A. . E. Bobick, "The representation and recognition of human movement using temporal template," in *In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1997.

- [73] M. J. Black and Y. Yacoob, "Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion," in *Proceedings of IEEE International Conference on Computer Vision*, 1995.
- [74] Z. Yin and R. Collins, "Moving object localization in thermal imagery," in *In: Proc. IEEE Workshop on Object*, NY, 2006.
- [75] I. A. Essa and A. P. Pentland, "Facial expression recognition using a dynamic model and motion energy," in *In Proceedings of IEEE International Conference on Computer Vision*, IEEE, 1995.
- [76] J. R. Bergen, P. J. Burt, R. Hingorani and S. Peleg, "A three-frame algorithm for estimating two-component image motion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(9), pp. 886-896, 1992.
- [77] Y. Kameda and M. Minoh, "A human motion estimation method using 3-successive video frames," in *In International conference on virtual systems and multimedia*, 1996.
- [78] R. T. Collins, A. J. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin and L. Wixson, "A system for video surveillance and monitoring," VSAM final report, 2000.
- [79] A. M. Husein, D. Halim and R. Leo, "Motion detect application with frame difference method on a surveillance camera," *In Journal of Physics: Conference Series*, pp. 1-11, 7 2019.
- [80] D. M. Tsai, W. Y. Chiu and M. H. Lee, "Optical flow-motion history image (OF-MHI) for action recognition," *Signal, Image and Video Processing*, pp. 1897-1906, 2015.
- [81] C. Shan, Y. Wei, X. Qiu and T. Tan, "Gesture recognition using temporal template based trajectories," *In Proceedings of the 17th International Conference on Pattern Recognition(ICPR 2004)*, pp. 954-957, 2004.
- [82] M. A. R. Ahad, T. Ogata, J. K. Tan, H. Kim and S. Ishikawa, "A complex motion recognition technique employing directional motion templates," *International Journal of Innovative Computing, Information and Control*, pp. 1943-1954, 2008.

- [83] R. Rosales and S. Sclaroff, "3D trajectory recovery for tracking multiple objects and trajectory guided recognition of actions," in *1999 IEEE computer society conference on computer vision and pattern recognition*, IEEE, 1999.
- [84] J. J. Wang and S. Singh, "Video analysis of human dynamics—a survey," *Real-time imaging*, pp. 321-346, 2003.
- [85] K. Alahari and C. V. Jawahar, "Discriminative actions for recognising events," In *Computer Vision, Graphics and Image Processing*, pp. 552-563, 2006.
- [86] J. K. Tan and S. Ishikawa, "High accuracy and real time recognition of human activities," in *IECON 2007-33rd Annual Conference of the IEEE Industrial Electronics Society*, IEEE, 2007.
- [87] C. Orrite, F. Martínez, E. Herrero, H. Ragheb and S. Velastin, "Independent viewpoint silhouette-based human action modelling and recognition," in *The 1st International Workshop on Machine Learning for Vision-based Motion Analysis-MLVMA'08*, 2008.
- [88] M. Vafadar and A. Behrad, "Human hand gesture recognition using motion orientation histogram for interaction of handicapped persons with computer," in *International Conference on Image and Signal Processing*, Berlin, Heidelberg, 2008.
- [89] W. C. Yau, D. K. Kumar, S. P. Arjunan and S. Kumar, "Visual speech recognition using image moments and multiresolution wavelet images," in *International Conference on Computer Graphics, Imaging and Visualisation (CGIV'06)*, 2006.
- [90] W. C. Yau, D. K. Kumar and S. P. Arjunan, "Voiceless speech recognition using dynamic visual speech features," in *In Proceedings of the HCSNet workshop on Use of vision in human-computer interaction-Volume 56*, 2006.
- [91] J. Han and B. Bhanu, "Individual recognition using gait energy image," *EEE transactions on pattern analysis and machine intelligence*, 28(2), pp. 316-322, 2006.

- [92] J. Liu and N. Zheng, "Gait history image: a novel temporal template for gait recognition," in *2007 IEEE international conference on multimedia and expo*, 2007.
- [93] Q. Ma, S. Wang, D. Nie and J. Qiu, "Recognizing humans based on gait moment image," in *Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD 2007)*, 2007.
- [94] Q. Ma, S. Wang, D. Nie and J. Qiu, "Gait recognition at a distance based on energy deviation image," in *2007 1st International Conference on Bioinformatics and Biomedical Engineering*, 2007.
- [95] Z. Liu, "Gait recognition using Active Energy Image and Gabor wavelet," in *2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 2016.
- [96] G. R. Bradski and J. . W. Davis, "Motion segmentation and pose recognition with motion history gradients," *Machine Vision and Applications*, 13(3), pp. 174-184, 2002.
- [97] M. Pantic, I. Patras and M. F. Valstar, "Learning spatio-temporal models of facial expressions," in *Int'l Conf. Measuring Behaviour*, 2005.
- [98] A. R. Ahad, J. K. Tan, H. S. Kim and S. Ishikawa, "Analysis of Motion Self-Occlusion Problem Due to Motion Overwriting for Human Activity Recognition," *Journal of Multimedia*, 5(1), pp. 36-46, 2 2010.
- [99] T. H. Lam and R. S. Lee, "A new representation for human gait recognition: Motion silhouettes image (MSI)," in *International conference on biometrics*, Berlin, Heidelberg, 2006.
- [100] S. Chen, W. Huang, Q. Guo and L. Dong, "Wavelet moments for gait recognition represented by motion templates," in *2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery*, 2010.
- [101] H. Foroughi, A. Naseri, A. Saberi and H. S. Yazdi, "An Eigenspace-Based Approach for Human Fall Detection using Integrated Time Motion Image and

- Neural Network," in *2008 9th International Conference on Signal Processing*, 2008.
- [102] T. Dobhal, V. Shitole, G. Thomas and G. Navada, "Human Activity Recognition using Binary Motion Image and Deep Learning," *Procedia Computer Science*, pp. 178-185, 2015.
- [103] R. Igual, C. Medrano and I. Plaza, "Challenges, issues and trends in fall detection systems," *Biomedical engineering online*, pp. 1-24, 9 2013.
- [104] X. Wang, J. Ellul and G. Azzopardi, "Elderly fall detection systems: A literature survey," *Frontiers in Robotics and AI*, 7, p. 71, 23 7 2020.
- [105] J. Gutiérrez, V. Rodríguez and S. Martin, "Comprehensive Review of Vision-Based Fall Detection Systems," *Sensors*, 21(3), pp. 1-50, 2021.
- [106] "An automatic in- ,P. Cuddihy و M. Rantz ,M. Skubic ,M. Popescu ,L. Liu
ournal of "home fall detection system using Doppler radar signatures
.2016 ,pp. 453-466 ,*Ambient Intelligence and Smart Environments*, 8
- [107] W. Min, H. Cui, H. Rao, Z. Li and L. Yao, "Detection of human falls on furniture using scene analysis based on deep learning and activity characteristics," *IEEE Access*, 6, no. 6, pp. 9324-9335, 2018.
- [108] S. Sherin P and Anju.J, "Human Fall Detection using Convolutional Neural Network," *International Journal of Engineering Research & Technology (IJERT)*, pp. 1368-1372, 7 2018.
- [109] M. Salimi, J. J. M. Machado and J. M. R. S. Tavares, "Using deep neural networks for human fall detection based on pose estimation," *Sensors*, p. 4544, 2022.
- [110] S. Li, X. Song, S. Xu, H. Qi and Y. Xue, "Dilated spatial-temporal convolutional auto-encoders for human fall detection in surveillance videos," *ICT Express*, p. 7, 2022.

- [111] Y. ZHU و W. YAN , "Ski Fall Detection from Digital Images Using Deep Learning " , *ACM ICCCV* , p. 13 .2022 ،
- [112] T. CHEN, Z. DING and B. LI, "Elderly Fall Detection Based on Improved YOLOv5s Network," *IEEE Access*, pp. 91273-91282, 2022.
- [113] F. Kamel and A. Salem, "Fall detection for elderly-people monitoring using Deep Learning," *UNIVERSITY OF ECHAHID HAMMA LAKHDAR - EL OUED*, p. 66, 2022.
- [114] C. Yao, J. Hu, W. Min, Z. Deng, S. Zou and W. Min, "A novel real-time fall detection method based on head segmentation and convolutional neural network," *Journal of Real-Time Image Processing*, pp. 1939-1949, 6 5 2020.
- [115] A. Tözeren, "Human Body Dynamics:Classical Mechanics and Human Movement," in *Human Body Dynamics:Classical Mechanics and Human Movement*, Washington, DC 20064, USA, Springer, 2001, p. 21.

Abstract

The prevalence of artificial intelligence techniques in the last decade, and the widespread of surveillance cameras, have contributed to the classification of human motions. Especially those that may have a negative impact on their health, and may, in worst cases, lead to their death. So, in this thesis two fall detection systems have been proposed. The first proposes a development of hybrid network CNN-LSTM with new feature vector connecting between CNN and LSTM. This new vector has general and special features, and recognizes three cascaded staging positions. These positions recognize different types of fall. The resulted accuracy is 98.6% and the sensitivity is 100%. This proposed method is superior to others that used the hybrid network CNN-LSTM. This method is trained on standard datasets (FDD, URFD, U. Montreal).

The second one introduces a new method, called STTFI, that extracts a person's silhouette with spatial and temporal characteristics representing the motions and its quantities and orientation such as previous methods (MHI, BMI, ITMI) that have drawbacks. The first one is that these methods are unable to represent the static parts of body which are either pivots or transient static poses during the activity. The second one is self-occlusion. But STTFI overcomes these drawbacks. Then it has been used with a low-depth convolution neural network (Alexnet) using learning transfer technique to classify the event as fall or not fall. The resulted accuracy is 99.02% and the sensitivity is 100%. This proposed method is superior to ones that used MHI with CNN network, multi-stream network or hybrid network CNN-LSTM. This method is trained on standard datasets (FDD, URFD, U. Montreal) and FD22 dataset that proposed for this research. FD22 contains images that have never been in standard datasets like fall of fainting and falling down from sofa.

Keywords: Fall detection, Elderly people, Motion History Image (MHI), Convolution Neural Network (CNN), Deep learning, Transfer learning, Computer vision, Integrated time motion images (ITMI), and LSTM cell.

Syrian Arab Republic
Damascus University
Faculty of Mechanical & Electrical Engineering
Department of Computer and Automation Engineering



Designing A System for Detecting the Harmful Fall of Human Based on Surveillance Cameras

A dissertation submitted in fulfilment of the requirements for the degree of PhD in Control
and Automation Engineering

By

Eng. Ammar MHD Yasser Aldaqqaq

Supervised by

Dr. Mohamed Mazen Almahayri

Dr. Rasha Kadem Massoud

Academic Year

2023