Title	Gait recognition from ill-posed silhouettes
Author(s)	Uddin, Md. Zasim
Citation	大阪大学, 2020, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/76644
rights	
Note	

論 文 内 容 の 要 旨

氏　名　（	Md. Zasim Uddin　）

論文題名	Gait recognition from ill-posed silhouettes (不良設定シルエットからの歩容認証)

論文内容の要旨

The importance of automated visual surveillance in a public space has been increased in response to the recent rising concerns about a safe and secure society. Gait, which can be defined as the manner of walking of a person, has been recognized as a potential biometrics modality, and it enables person recognition from a low-resolution image sequence in a CCTV footage. Image sequence can be captured at a distance without the subject's cooperation, and information associated with shape and his/her walking motion can be obtained from the binary silhouettes from the captured image sequence. Therefore, person recognition using silhouette-based gait features (i.e., gait recognition) is a promising real-life application. However, practical deployment in a real-world scenario is still considered a challenging task because several issues should be managed to achieve efficient gait recognition. Such issues can be related to the presence of external factor to a subject (e.g., carried object, clothing, shoes); the internal factor to a subject (e.g., walking speed, aging, pregnancy); the environmental context (e.g., illumination, walking surface, occlusion) and the quality of captured image sequence from a camera (e.g., spatial resolution, temporal resolution, observation view). These factors make the gait recognition more challenging and prevent the use of gait recognition in the real-world scenario. In order to realize efficient gait recognition in the situation with those factors, this thesis addresses the following three issues: i) carried object(s) (COs) ii) occlusion and iii) data quality.

The factors are considered in this thesis, affect the captured image sequence of a person for gait recognition. Therefore, the obtained silhouettes from the corresponding captured images can be either ill-posed or well-posed based on visual quality. The well-posed silhouettes mean that the obtained silhouettes are of good quality, whereas ill-posed silhouettes are degraded, problematic, and low-quality silhouettes. For example, COs silhouette is added to the person's silhouette, whereas some parts of a person's silhouette are unobservable due to the occlusion. Regarding data quality of the captured image sequence from a CCTV camera, the obtained silhouettes are different depending on several camera settings such as spatial resolution, temporal resolution, and observed view of a person from the camera.

First, in order to tackle the COs issue, we constructed the world's largest gait database with real-life COs and analyzed the performance of silhouette-based gait recognition. Whereas existing databases for gait recognition include at most 306 subjects with COs, we constructed an extremely large-scale gait database that includes 62,528 subjects, with an equal distribution of males and females, and ages ranging from 2 to 95 years old. Moreover, existing gait databases include person image sequences with COs, but the variation of the COs' type, positions, and numbers are limited and predefined. However, our constructed database considered a wide variety of COs' types, positions, and numbers. Besides, the silhouettes data in the database were manually annotated into seven carrying status (CS) labels. The extremely large-scale gait database with CS labels enabled us to evaluate and analyze the impact of the training data size, the recognition difficulty level of the CS labels, and the classification of CS labels.

Research for the second issue is managing occlusion. The presence of occlusion in an image sequence obscures a significant amount of the human body. However, usual gait recognition methods require an unoccluded silhouette sequence. Thus, occlusion handling is an important issue for gait recognition. We, therefore, propose a silhouette sequence reconstruction approach from an occluded sequence (sVideo) based on leveraging the powerful capabilities of conditional deep generative adversarial network (GAN). To reconstruct the occluded silhouette sequence well constrained, we regularize the training of the proposed generative network, based on triplet hinge loss incorporating with Wasserstein GAN (WGAN-hinge). To the best of our knowledge, WGAN-hinge is the first adversarial loss that supervises the generator network during training by incorporating pairwise similarity ranking information. The proposed approach was evaluated on multiple challenging occlusion patterns. The experimental results

demonstrate that the proposed approach outperforms existing state-of-the-art benchmarks.

Finally, we considered the quality of the captured image sequence from a CCTV camera; particularly, we analyze the camera-based qualities of the spatial resolution (SR) and temporal resolution (TR) along with the observation view because they affect the recognition accuracy. It is well-known that uni-modal biometrics is usually affected by some practical problems such as poor recognition performance and vulnerability to spoofing attacks. Therefore, we extracted gait, head, and soft biometric height modalities from the captured image sequence obtained from a single camera and introduce SR, TR, and view as the quality that significantly affects biometric system performance. Moreover, data quality (i.e., quality measure) that affects the matching score of each modality is often incorporated as a quality-dependent score-level fusion, which is a popular and promising approach. We considered seven and ten scaling factors for SR and TR, respectively, with four view variations. Therefore, a large database is constructed, comprising approximately four million genuine and 7.5 billion imposter scores. Later, we analyze the recognition accuracies associated with gait, head, and height modalities in different spatial and temporal resolutions along with the observation views. We observed that TR impacts profoundly on the gait modality, whereas SR for the head modality. These analyses can be useful for forensic analysis, mainly when the quality of the captured image sequences is very poor. Finally, we evaluate and analyze the performance considering all of the modalities in a score-level fusion by designing two different protocols that are the quality-independent and quality-dependent manner using state-of-the-art score-level fusion approaches. These evaluation results will be beneficial for score-level fusion research.

This thesis together with the considered issues could make a significant contribution to improving the gait recognition performance in various real-world scenarios that further enable the adoption of gait-based person recognition into automated visual subservience and forensic applications.

論 文 審 査 の 結 果 の 要 旨 及 び 担 当 者

氏　　名	（ Md. Zasim Uddin ）		
論文審査担当者	（職）		氏　　名
	主 査　　　　　教授		八木　康史
	副 査　　　　　教授		楠本　真二
	副 査　　　　　教授		長原　　一（大阪大学データビリティフロンティア機構）
	副 査　　　　　准教授		村松　大吾

論文審査の結果の要旨

　本論文で、申請者は生体認証の一つである歩容認証において、ill-posed silhouette から個人を認証する研究成果についてまとめている。現在歩容認証では、シルエットに基づく認証手法が主流となっており、高品質の、きれいな人物シルエットを抽出することができれば、高い精度での認証が可能となっている。しかしながら、実際の歩容認証の適用場面を想定すると、高品質の、きれいな人物シルエットを抽出できることは稀である。例えば、荷物を所持している人物のシルエットは、人物シルエットに加え、所持している荷物のシルエットが付加されたものになるし、物などの遮蔽物がある場合には、遮蔽された領域の情報は取得することができず、抽出される人物シルエットは、遮蔽部分が欠けたものとなってしまう。つまり、抽出される人物シルエットは低品質の、劣化したものとなってしまうことが多い。本論文は、このような、低品質の劣化した人物シルエットをill-posed silhouettesとし、ill-posed silhouettesから歩容認証を実現するアプローチについてまとめている。具体的な内容としては、本論文では、荷物所持に基づく人物シルエット劣化の問題、隠ぺいによる人物シルエット劣化の問題、カメラ設定等に基づく人物シルエットの品質低下の問題に焦点を当て、それぞれにおいて貢献を報告している。

　荷物所持に基づく人物シルエット劣化の問題においては、最新の機械学習アルゴリズムに基づく識別器を対象としているが、識別器を有効に学習するためには、適切なデータベースが必要である。しかしながら、既存の荷物所持に係る歩容データベースは、被験者のバリエーション及び荷物所持方法のバリエーションの二つの点で不十分であり、そのようなデータベースからの学習では、実環境での歩容認証を高精度で実現する識別器学習は困難であった。そこで、申請者は、6万人を超える被験者から、自由な荷物所持方法の歩容データを収集し、世界最大の歩容データベースを構築することで、荷物所持歩容認証の高精度化に貢献している。

　隠ぺいによる人物シルエット劣化の問題においては、欠損のある人物シルエット系列から、欠損のない歩容特徴を再構成できる手法を構築している。この構築手法においては、深層学習における敵対的生成ネットワークの一手法である、Wasserstein GAN（WGAN）を用い、特徴再構成の際の再構成誤差に加え、同じ人物の再構成歩容特徴は近づくよう、また別人物の再構成歩容特徴は遠ざかるような損失関数を設定し学習することで、高精度化を実現可能な手法を構築している。構築手法は、様々な隠ぺい状況を想定した評価実験において他手法よりも高い認証精度を実現しており、隠ぺい状況下での歩容認証実現に大きな貢献をしている。

　カメラ設定等に基づく人物シルエットの品質低下の問題においては、認証対象となる歩行映像から取得可能な歩容特徴以外の頭部特徴、身長特徴も抽出し、人物シルエットの時間解像度、空間解像度等を品質の定量指標として、品質に応じて、歩容、頭部、身長を適切に組み合わせるマルチモーダル手法を提案し、評価を行っている。実環境においては、シーンに応じてシルエットの品質が異なることが一般的であるため、この手法の有効性は高い。

　このように、本論文では実環境における高精度人物認証を実現するために非常に重要である、ill-posed silhouettesからの歩容認証を研究対象としており、その対象において、情報科学的なアプローチによる重要な貢献がまとめられている。

よって、博士（情報科学）の学位論文として価値のあるものと認める。