



Title	Learn to Walk Across Ages: Cross-age Gait Analysis with Spatio-temporally Augmented Representation
Author(s)	張, 億一
Citation	大阪大学, 2022, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/91776
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

Abstract of Thesis

Name (Yiyi Zhang)	
Title	Learn to Walk Across Ages: Cross-age Gait Analysis with Spatio-temporally Augmented Representation (時空間的な拡張表現による歩容の年代間解析に関する研究)
<p>Abstract of Thesis</p> <p>Video is becoming an important carrier of information in this digital era under the development of communication technology and widespread availability, where surveillance video is one of the most important types. Due to the widely deployed surveillance system, large-scale videos are captured everyday, making the computer-aided surveillance analysis system necessary.</p> <p>Gait is a very important subject in surveillance video due to its noncontact, noninvasive and can be perceivable at a long distance. Psychological experiments have shown that gait includes not only identity information, but is also discriminative in a variety of human attributes such as age and gender. Among all the human attributes, since age is inevitable in person's life span, it is one of the crucial factors in gait analysis.</p> <p>In this thesis, we propose the task of cross-age gait video translation, which aims to realize a function of age progression/regression on gait. The translated gait videos attempted to preserve the identity of the original subject while keeping the realism of the generation quality. Cross-age gait video translation can not only improve the fundamental understanding of age features in a subject, but also support a range of applications such as criminal investigation and fugitive research. Moreover, cross-age gait translation can be a promising solution for age-invariant gait recognition.</p> <p>In Chapter 1, we first review the background and motivation of the proposed cross-age gait video translation task. we then summarize two aspects in realizing this: gait video translator with spatio-temporally augmented network design, and the spatio-temporally augmented high fidelity input.</p> <p>In Chapter 2, to learn an effective gait video translator across ages, we proposed spatio-temporally augmented multi-age group gait video translation framework, which aims to ensure three aspects: aging effect, individuality preservation, and gait realism. Specifically, we build our framework on a multi-domain image translation model. Because the existing multi-domain image translation model was originally designed for a still image, we extend it to gait video by introducing a motion-augmented network architecture with three streams, where gait period, period-normalized phase-synchronized gait video, and its frame difference sequence are each input to one stream. We also designed a discriminator with a slow-fast path to learn spatio-temporally augmented gait representation for cross-age gait video translation. Our framework quantitatively and qualitatively outperforms state-of-the-art age progression/regression methods on the largest gait database with age, OULP-Age, with respect to both age group classification and identity recognition.</p> <p>In Chapter 3, since appearance-based gait analysis approaches usually use silhouette or silhouette-based template as input, silhouette quality plays an important role in gait analysis. To learn a spatio-temporally augmented high fidelity input for the proposed cross-age gait video translation task, we studied natural image matting task and achieved competitive results on widely acknowledged matting benchmarks. To capture global contextual information from a whole image without degrading the image quality, an end-to-end three-branch image matting framework is proposed, which can exploit unknown-relevant global contextual information condensed from the high-resolution image. We then proposed a matting-oriented contextual aggregation can cope with such a situation by making use of all the pixels in the deformed foreground/background where foreground/background pixels are dominant. The proposed method can estimate alpha matte and background simultaneously while keeping the matting equation, which can improve the foreground extraction performance qualitatively.</p> <p>In Chapter 4, We further provided a discussion on how the high fidelity input may influence the gait video translator across ages. We first designed a scheme to automatically estimate the trimap for the proposed matting method, so that the matting can be adopted without user interaction. Specifically, we adopted a inpainting method to predict the background and finetuned the proposed matting method on OULP-</p>	

Age. We then provided thorough experiments on the largest gait database with age information, OULP-Age, to reveal how the input quality of silhouette affect the performance of age progression/regression task on age group classification and cross-age gait recognition.

Finally, conclusions are drawn and future work is discussed in Chapter 5.

論文審査の結果の要旨及び担当者

氏 名 (Yiyi Zhang)			
論文審査担当者	(職)	氏 名	
	主 査	教授	八木 康史
	副 査	教授	槇原 靖
	副 査	教授	長原 一
	副 査	准教授	中村 友哉

論文審査の結果の要旨

本論文は、歩容（歩き方の個性）に基づく個人認証技術を長期間に渡る科学捜査等へ適用することを目的として、時空間的な拡張表現による歩容の年代間解析手法を提案したものである。

第一章では、人物映像解析の一分野である歩行映像解析が、科学捜査やセキュリティの分野において重要な役割を果たすことを論じている。一方で、人の歩容は、子供の成長期や高齢者の運動機能低下期等、経年変化の影響を受けやすいことから、長期に渡る犯罪捜査や失踪人物捜査に歩容認証を適用する際には、経年変化への対応が必要となる。そこで、本研究では、経年変化に頑健な歩容認証に向けて、歩行映像のある年代から別の年代へと個人性を保ったまま変換するモデルを提案している。特に、変換モデルの入力となる歩行映像の品質、並びに、変換モデルのネットワーク構造の二つが重要であると考え、その各々に対して、効果的な時空間表現を導入している。

第二章では、歩行映像の年代間の変換モデルを提案している。変換モデルの入力には、体型等の空間的情報を表す歩容シルエット映像に、歩容の動きや速さ等の時間的情報を表すフレーム間差分映像や歩行周期（左右計2歩分に要する時間）を加えた、合計三つを用いる。また、変換モデルの基幹ネットワークにおいて、上記の空間・時間的情報を相互に有効活用するための Motion Augmented Block や SlowFast Path を提案している。実験では、歩行映像の公開データベース OUMVLP-Age の実データを元にして、様々な年代の歩行映像を生成することで、異なる年代間での歩容認証のシミュレーション評価を行っている。結果として、他の年代変換モデルと比較して、定性的にも優れた歩行映像の変換結果が得られており、個人認証及び年代識別の精度面でも有効性を示している。

第三章では、背景を0、前景（人物領域）を1として、その中間値を表現可能なアルファマットを推定する手法を提案している。アルファマットにより、空間的側面である人物の境界形状が滑らかに表現され、また、時間的側面である動きぼけも表現されることから、背景・前景の二値からなるシルエット画像と比較して、アルファマットは時空間的に拡張された入力表現であると論じている。また、アルファマットを推定する際には、前景・背景の情報を抽出する必要があるが、深層学習モデルのパラメータ数の制約により、原画像と比べて低い空間解像度の画像からサンプリングすることとなる。そこで、その限られた空間解像度を効果的に活用するため、前景／背景が画像中で支配的になるように空間的に変形した画像からサンプリングする Deformable sampling 手法を導入している。実験では、アルファマットの公開データセットを用いて、定性評価並びに定量評価を実施している。結果として、提案手法により、従来手法よりも高精度なアルファマットが得られることを示している。

第四章では、第三章の手法で推定されるアルファマットを、第二章の歩容の年代間変換モデルに入力することで、提案する枠組み全体としての個人認証の精度評価を実施している。結果として、二値のシルエット映像を入力する場合と比較して、より高い個人認証精度が得られることが示された。

第五章では、本研究を総括し、今後の展望について述べている。

以上により、本論文は博士（情報科学）の学位論文として価値あるものと認める。