

学位論文 博士（工学）

被写体の姿勢変動によらない顔認識システム  
に向けた顔画像の時空間解析

2011 年度

慶應義塾大学大学院 理工学研究科

田中 秀典

# 主 論 文 要 旨

報告番号	甲 乙 第	号	氏 名	田中 秀典
主 論 文 題 目： 被写体の姿勢変動によらない顔認識システムに向けた顔画像の時空間解析				
(内容の要旨) 顔認識技術は、コンピュータビジョン分野で古くから研究されており、近年においてはマンションの入退室管理やデジタルカメラへの応用など実用化が進んでいる。しかしながら、これまでの顔認識技術では、被写体がカメラに正対し、被写体の正面顔画像が得られることが前提となっており、被写体の姿勢に依存しない顔認識は未だに困難な問題として研究開発されている。 本論文では、被写体の姿勢に制限を与えずに撮影された画像から、時空間解析に基づいて顔認識するための新しい手法を提案し、その有効性を実験により検証する。 本論文では、まず、近年多くの場所に設置されている複数台の監視カメラから撮影された複数枚の画像を利用することにより、被写体の姿勢に制限を与えない顔認識手法を提案する。本手法においては、異なる視点のカメラから撮影された複数の被写体の顔画像を、頭部の標準3次元形状に貼り付け、任意姿勢の被写体の顔画像を生成することにより姿勢変動への対処を行う。この際、顔画像生成の元となる被写体の顔画像が少数であると、被写体の様々な姿勢変動に対応し高い精度の識別を行うことが困難になるので、識別した顔画像を追加学習することで情報量をふやす。ただし、得られた顔画像の姿勢が、すでに学習を行った顔画像の姿勢近辺であれば、追加で学習を行わないようにする。 続いて本論文では、ロボット中のカメラ等の位置・姿勢を制御可能なカメラを利用することにより被写体の姿勢に制限を与えず、さらに対話（インタラクション）時に得られる動画画像から表情変化の個人性を利用して、顔認識を行う手法を提案する。本手法においては、表情変化による各顔部分の変化過程を特徴量として表現する。 前者の多視点画像を利用した顔認識手法は、空間の3次元情報解析に基づくものであり、後者の動画画像を利用した顔認識手法は、画像の時間解析に基づくものである。これらを総称して本論文では時空間画像解析と位置付ける。 提案手法の有効性を確認するため行った実験より、環境中に複数カメラが設置されている場合、初期学習時4枚の顔画像と標準3次元形状モデルで生成した画像を用い、追加学習および選択学習を行うことにより、被写体の姿勢変動によらず顔認識が高い精度で行えることを示した。また、インタラクションにより生じた表情変化を、提案手法により各顔部分の変化過程として表現することで、表情変化前後の各顔部分の動きの差として表現するよりも、顔認識が高い精度で行えることを示した。特にhappyの表情において顕著な精度向上が見られ等価エラー率が約4%減少した。 これらの結果により、本論文で提案した時空間画像解析手法は被写体の姿勢変動によらない顔認識に有効であることが確認できた。本提案手法によって実現可能となる顔認識システムは、その場・その時に応じた情報提供サービス（例えば、経験の浅い看護師への注意喚起、製造現場作業者の作業フロー学習、買い物レコメンド）の実現へつながっていくと考えられる。				

## SUMMARY OF Ph.D. DISSERTATION

School Science and Technology	Student Identification Number	SURNAME, First name TANAKA Hidenori
Title  Spatio-temporal analysis of facial images for pose invariant face recognition system		
Abstract  <p>This thesis presents spatio-temporal analysis of facial images for pose invariant face recognition system. Face recognition has been studied in the field of computer vision for about 50 years. In recent years, face recognition techniques from frontal view images are utilized in access control system and digital cameras. One of the major challenges by current face recognition techniques lies in the difficulties of handling various poses.</p> <p>Sparsely distributed surveillance cameras capture people from various views. However, it is not possible to identify people from various viewing angles because the various appearances cannot be effectively learned with sparsely distributed images. In our approach, virtual viewpoint images are synthesized by interpolating the sparsely distributed images with a simple 3D shape of the human head, so that virtual densely distributed images can be obtained. These synthesized images enable to identify people from various viewing angles. An initial eigenspace is generated using these synthesized images in the initial learning step. In the following learning step, additional images captured by other distributed cameras are used for the eigenspace updates to improve recognition performance. However, this learning step is not performed, when the pose parameters of an additional image are similar to the pose parameters of the learned images.</p> <p>Communicative moving cameras (e.g. robot eyes) can capture facial expression changes in videos from a frontal or near-frontal facial view while communicating with people. In our approach, dynamics of facial expression changes in videos are analyzed to identify people.</p> <p>Analysis of sparsely distributed images is a sort of spatial analysis, and analysis of facial expression changes in videos is a sort of temporal analysis. In this thesis, we call these analyses a spatio-temporal analysis.</p> <p>Experimental results showed that the high discernment rate is achieved by repeating the additional learning step when the four sparsely distributed images are interpolated with a simple 3D shape of the human head in the initial learning step. We also found that facial expression changes represented by the proposed method have higher discriminating power than that represented by the previous method. Particularly, in the case of a happy face, the proposed method reduced about 4% of EER compared with the previous method.</p> <p>These sophisticated methods are available for pose invariant face recognition system. This system realizes personal service applications (e.g. reminder system for inexperienced nurses, learning system for untrained operators, shopping recommendation system).</p>		