論文内容の要旨

論文題目 Human Body/Head Orientation Estimation and Crowd Motion Flow Analysis from a Single Camera

(単一カメラ映像からの人の体と頭部の向きの推定と群衆の動き場の解析)

氏名　オストルク　オゥグ

In the last few decades, automation of **descriptive and statistical analysis of human behavior** became a very significant research topic. Due to the technological advances in video technologies, many researchers have focused on detection and analysis of human motion from video cameras. To achieve this, until now researchers have tried to solve problems, such as: detecting humans in a given scene, counting number of humans, motion tracking and analyzing their trajectory of motion, etc. However, to truly understand human behavior and evaluate a given scene, more semantic analysis is required. Hence, as a next step, **measuring people's visual focus of attention has become a significant and challenging problem**. Measuring focus of attention can be useful in many different ways. Finding the objects or places in people's focus of attention can help us to understand their intention and control the security of the environment. People looking at the bulletin boards, customers walking around market stands can provide information about recent trends, marketing strategies. Social interactions can be interpreted in a more meaningful way, human-computer interactions improve and more intelligent autonomous systems can be built.

Visual focus of attention of humans is defined as the direction they are heading to or the direction they are looking at during their motion. Humans show their attention by walking towards that direction or by turning their head to that direction. In a given scene, the orientation of body and head of a human can give us a hint about his/her visual focus of

attention. When there are many people in the scene, e.g. extremely crowded scenes, it is not possible to analyze each person individually. In this case, the paths humans mostly walk can give us information about their interests in the environment.

Currently, there are a huge number of researches that try to solve visual focus of attention estimation problem by using multiple cameras, multi-sensors or they locate various markers on the bodies. These approaches are often too impractical or expensive to build in common public places for general cases. Our aim is to extract the most possible useful information to achieve human motion analysis in a given public scene from a single camera. However, it has very big challenges due to the articulations in human pose and less data. On the other hand, **by only using a single camera, we can build portable, low-cost systems with less complexity**.

In our research we focus on two major problems. First, we have developed a system that tracks people and estimates their body and head orientation. Second, we have analyzed various crowd scenes and proposed a method to calculate the dominant motion flows that can handle very complex situations. At the beginning of our research, we have studied human tracking and developed a real-time application which tracks multiple people simultaneously and detects their focus of attention by estimating their next steps. It was incorporated in a digital art project that was exhibited in Haneda Airport in Tokyo for one month.

First of all, we have developed an interactive entertainment system employing real-time multiple human tracking to demonstrate possible applications and importance of human tracking and motion understanding. The system was a part of a digital public art project, which presented technological advances by using art in an airport. Our system tracks people and continuously visualizes their predicted future footsteps in front of them while they keep moving. A real-time multiple human tracking algorithm has been developed and combined with a visualization process. The system can be installed in any indoor place easily. It does not affect the natural flow of life in the sense that it does not affect the movements of people until they notice the displayed interactive foot shapes. Hundreds of passers-by visited the system for duration of one month. When they noticed, people showed surprise,

excitement, astonishment. They tried to discover where and why the foot shapes were coming from. They played with the system by making various movements.

To solve focus of attention analysis problem for a single person, we have developed a system which tracks a person in the environment and estimates his/her body and head orientation during their motion from a top-view single camera. We have utilized the edge map of Ω shape of the head-shoulder region of the person to estimate the initial body orientation. Next, by calculating the orientation change in the body and head, we estimated the new orientation. Displacement vectors of SIFT features have been analyzed to calculate the orientation change. The experiments showed that, body and head orientation of a person can be estimated successfully by using the proposed method. The orientation angle range was pi/8 and the error was five degrees at most. The algorithm works for a single person under various motions(walking straight, turning the head to the right, turning around himself, turning right, etc…) of the person in the scene.

Next, complex crowd motions have been studied to determine the focus of attention of crowds in the scenes. To analyze the crowd motion, one of the most useful information is to find the mostly followed paths. It gives us information about the tendency of the people in the scene and usage statistics of the regions in the scene. We have proposed two main algorithms to detect the dominant motion flows in structured and unstructured (very complex) scenes. First algorithm was developed to extract and represent the motion flows in the scene in the local regions. We used SIFT motion flows for short periods of time and we have accumulated those instantaneous flows for a long period to represent the motion in the scene. Then, we utilized a hierarchical clustering algorithm to classify the motion flows into meaningful groups by prioritizing the orientation. The proposed system was tested against a group of challenging scenarios from real world scenes and it successfully detected the dominant motion flows. Furthermore, our system provided a flexible way to analyze the motion flows in various levels of detail and it also successfully dealt with the local irregularities and detected the motion flows in any part of the scene.