Human Tracking using Multiple-Camera-Based Global Color Model in Intelligent Space

TaeSeok Jin¹, and Hideki Hashimoto²

¹Dept. of Mechatronics Engineering, DongSeo University, 
San 69-1 Churye-2 Dong, Sasang-Gu, Busan 617-716, Korea
²Institute of Industrial Science, University of Tokyo,
4-6-1 Komaba Meguro-ku Tokyo 153-8505, Japan

Abstract

We propose an global color model based method for tracking motions of multiple human using a networked multiple-camera system in intelligent space as a human-robot coexistent system. An intelligent space is a space where many intelligent devices, such as computers and sensors (color CCD cameras for example), are distributed. Human beings can be a part of intelligent space as well. One of the main goals of intelligent space is to assist humans and to do different services for them. In order to be capable of doing that, intelligent space must be able to do different human related tasks. One of them is to identify and track multiple objects seamlessly. In the environment where many camera modules are distributed on network, it is important to identify object in order to track it, because different cameras may be needed as object moves throughout the space and intelligent space should determine the appropriate one. This paper describes appearance based unknown object tracking with the distributed vision system in intelligent space. First, we discuss how object color information is obtained and how the color appearance based model is constructed from this data. Then, we discuss the global color model based on the local color information. The process of learning within global model and the experimental results are also presented.

Key Words: Distributed cameras, Object tracking, Color histogram, Global model, Intelligent environment.

1. Introduction

An intensive research has been going on in the field of intelligent space in recent years [1][2]. An intelligent space is the space where many intelligent devices, such as computers and sensors, are distributed. The environment is said to be intelligent, because of the cooperation of many intelligent devices. We proposed "Intelligent Space (iSpace)" [3][4] in order to achieve a human-centered services by accelerating the physical and psychological interaction between humans and environments. Color CCD cameras with processing and networking capability, are distributed in the iSpace. Therefore they are called "DINDs (Distributed Intelligent Network Devices)". The construction of Intelligent Space is shown in Fig.1. DINDs record the position and behavior of both humans and robots in the iSpace. DINDs are capable of mutual communication. The position of humans and mobile robots in iSpace based on the information gathered by DINDs has been studied before. The most common approaches are color marker tracking [5], human behavior recognition[6] and mobile robot control in iSpace[7].

The iSpace has to track the objects without failure and to get the location of objects by DINDs seamlessly for these services. Seamless tracking and localization of objects must be achieved in order that the iSpace works properly. In this paper, color appearance based object representation for the distributed vision system in the iSpace is described. First, distributed vision system in the iSpace will be explained. Next, color appearance based object tracking, which is currently possible in the iSpace only, will be proposed. Then, this paper will show how to obtain color information of the objects, and how to achieve the correspondence among different cameras by using the object global color models.

Fig. 1. concept of Intelligent Space

2. Multi-Vision system in Intelligent Space

2.1 Multi-Camera Multi-Object Tracking

We have to solve the multiple-camera multiple-object tracking problem for the wide area sensing in the iSpace. There
are two major problems concerning multiple camera multi-
ple-object tracking system. One is the traditional corre-
spondence problem from frame to frame over time. The other
is the correspondence problem among different camera mod-
ules in order to achieve seamless tracking and location
estimation. The later problem is called the consistent-labeling
problem[8]. There are several approaches to solve this prob-
lem in recent papers.

These approaches include feature matching[9], location
information[10], [11], and alignment approach[12]. If all cam-
eras are calibrated in advance, consistent labeling can be es-
blished by projecting the location of each object in the world
coordinate system. Alignment approaches rely on recov-
ering the geometric transformation between the cameras.
However, it is difficult for these approaches to establish con-
sistent label without overlapping of the monitoring areas
among different cameras.

Feature matching approaches whether based on the color or
not are the simplest scheme to establish consistent labeling.
However, color feature matching is not reliable when the dis-
parity is large in location and orientation. For example, if a
person is wearing a shirt that has different colors on front and
back, simple color matching among different cameras doesn't
work. On the other hand, color information is useful for rec-
ognition and identification of objects in the interpersonal
communication. If color representation, that absorbs the dif-
ferences among different cameras and includes the color appear-
ce model of all round the object, is achieved, color infor-
mation is also useful for object identification in the com-
munication among different camera modules.

2.2 Cooperation of DINDs

It is desired that each object model is effective for cooper-
ation of the DINDs. Here cooperation of the DINDs means the
processing that each DIND exchange the information about the
objects in each monitoring area for the other DIND. Then,
correspondence of the object among different DINDs must be
achieved by using the effective object model. Literature[10],
[11] are representative of researches aiming at objects tracking
by cooperation of multiple cameras.

The color histograms of an extracted objects are used for the
appearance-based identification. The object representation based
on the color histogram is relatively stable against deformation
and occlusion [13]. Compared with the contour and so on, color
histogram of the object stays largely unchanged against the vari-
ous images that are captured by the distributed cameras. The
object representation using color histogram is suitable for ap-
pearance-based correspondence of multiple objects seamlessly
in wide area. It is difficult to realize the correspondence among
different cameras using the simple color histogram measured
from one direction. On the other hand, color histogram repre-
senting the current appearance is needed for color region track-
ing in one camera image. Two kinds of object models are de-
finned as follows to satisfy these requirements.

• Local Color Model

Local Color Model is used for object tracking and object
segmentation under the occlusion in one camera image. It is
updated according to the appearance of the object every
frame.
3. Acquisition of object information

3.1 Object Finding Process

Object finding process is the process to find the new moving objects in the monitoring area of the camera module. The local color models of new objects are also acquired. Background subtraction is simple and efficient to find the new moving objects in fixed camera image. This background subtraction uses the background model updated from frame to frame adaptively. In Intelligent Space, this background subtraction works well since the lights and the floor are configured as reducing the effects of the shadow and lighting condition. The candidate regions of moving objects are extracted after the dilation, erosion, and clustering to the binary image separated from captured image by comparison with the background image. The small object region is removed as the noise.

The initial local color model is defined as follows. \((x_i) = 1, \ldots, n\) is the pixel locations in the region extracted as the object. The function \(b\) associates to the pixel at location \(x_i\) the index \(b(x_i)\) of its bin in the quantized feature space. Feature space is represented by two-dimensional normalized color space, e.g., \(r = \frac{R}{(R+G+B)}, g = \frac{G}{(R+G+B)}\). The component \(p_{x,u}, u = 1, \ldots, m\) of the feature vector \(p\) in the object is then computed as

\[
p_{x,u} = \frac{1}{n} \sum_{x \in X} \delta(b(x) - u)
\]

where \(\delta\) is the Kronecker delta function.

Since the region extracted by background subtraction is unstable, several sets of \(p\) are required for each object in order to stabilize the initial local color model. There is also a probability that multiple objects are found simultaneously. The set of \(p\) should be clustered to some categories by the online clustering algorithm. It is decided whether obtained feature vector \(p\) belongs to any existing clusters or a new cluster is generated. The number of existing cluster is \(N\) at that time. At first, the similarity between feature vector \(p\) and each reference vector \(r_k\) of cluster is calculated to decide nearest neighbor cluster by Eq.(2). \(p_{i,t}\) denotes \(j\)-th object at the current time \(t\).

\[
S(p_{i,t}, r_k) = \sum_{u=1}^{m} \min(p_{i,t,u}, r_{k,u})
\]

(2)

It is assumed that \(c\) represents the adequate cluster, and it is computed as

\[
c = \begin{cases} \arg \max_k S(p_{i,t}, r_k) & \text{for } S(p_{i,t}, r_k) > T \\ N+1 & \text{otherwise} \end{cases}
\]

(3)

where, \(T\) is the threshold to evaluate the similarity between feature vectors.

The reference vector of each cluster is updated by Eq.(4). Updated vector is used as the reference vector at the next time \(t+1\). \(a\) is the learning coefficient.

\[
r_{k,i+1} = r_{k,i} + a \delta \delta(p_{i,t} - r_{k,i})
\]

(4)

When the vectors beyond the threshold are gathered in one cluster, object candidate which corresponds with this cluster is treated as the target object. The tracking process for each target object runs at that time. The reference vector \(r_{k,i}\) of each cluster is treated as the local color model \(l_{k,i}\).
42

4. Global model for object identification

4.1 Global Color Model

The global model means an object model for matching of objects measured by the different camera modules as mentioned above. Global model learning process runs when the occlusion among the objects doesn't happen in the tracking process. The global model is produced from the local color model which has been measured since the object finding process started. This model should cancel the effects of the object posture, scaling or the direction of measurement by cameras for matching between the different camera modules. At first, it is decided whether the global model is learned, based on objects conditions such as overlapping and approaching. This condition is decided by evaluating the distance between the bounding boxes.

The local color model can be configured every frame in tracking process. For example, the back of tracked human is captured in Fig. 9. The set of the local model includes the color appearances that change according to the posture of the object. If the color appearances of all round the object are in-
Human Tracking using Multiple-Camera-Based Global Color Model in Intelligent Space

cuded in the object model, it is useful for consistent labeling from the view of different cameras. It is not reasonable to store all local color models to each object in terms of the memory size. Global color model requires the effective representation from all local color models. The theory of Eigenspace Method, which is excellent in compression of large amount of image data and calculation of the correlation among images, is reported in [17]. Eigenspace method is applied to configuration of the global color model from a lot of local color models. Fig. 8 shows the details of this process.

\[ \lambda_k e_i = Q e_i, \quad (\lambda_1, \ldots, \lambda_d) \]  

The \( d \)-dimensional subspace spanned by these \( d \) eigenvectors corresponding to \( d \) large eigenvalues is called the eigenspace. By ignoring the small eigenvalues, dimension of the local color model data is reduced. The cumulative proportion of eigenvalues in Eq.(8) is evaluated in order to determine the effective dimension.

\[ W_d = \sum_{i=1}^{d} \lambda_i \]  

Then, one local color model is projected onto the eigenspace by

\[ z_{k,t} = [e_1, \ldots, e_d]^T L_{k,t} \]

\( z_{k,t} \) is a point that the local model \( L_{k,t} \) at \( t \) is mapped to the eigenspace. The local models to each object can be represented as a manifold in the eigenspace. This manifold includes the local color models changing according to the posture of the object. Global color model \( g_k \) is represented as this manifold.

### 4.2 Global Model Learning Process

Figure 9 shows the relationship between the acquisition of object information, which includes the object finding process and tracking process, and the process for leaning the global color models. This global model learning process is also included in the DIND. As we have said above, local color models of each object are acquired in every frame that occlusion between objects isn't found. Because the local color models are updated depending on the appearance of the object,
changed color models can be obtained in each image frame. The set of these local color models is stored into the global color model learning process. Eigenvalues and eigenvectors for the set of the local models of each object are calculated individually in this process.

5. Experiments

We implemented the proposed method to represent walking human position, used a global color model to tracking the top positions of each person and implemented a MSTP proposed in section III. Also, tracking experiments in which our method coped with the sudden change of the object movement were performed. Figure 11 demonstrates a typical eigenvector of three walking humans in a convoy, where the global color models of position are shown.

![Fig. 11. global color models of three humans](image)

![Fig. 12. Human's walking routes](image)

![Fig. 13. Comparison of the Global Models](image)

Figure 12 shows the routes of three humans tracking, respectively. Objects, three humans, were tracked by the tracking process which was performed independently with the other process. In Fig. 13, the global color model obtained in the different camera is compared with the tracked humans as shown in Fig. 12. A lot of local color models are projected onto the eigenspace spanned by three eigenvectors.

The global model is represented as the data sets of the local models compressed in this case. The cumulative proportion of eigenvalues in Eq.(8) and the threshold generally set to 0.8 or 0.9 to determine the effective dimension. This model can represent the change of the color appearance of the tracked human. Although the local color models change according to the difference of the camera, correspondence among cameras can be evaluated by the comparison of the manifold shape. The global models by camera2 and camera3 are in process of the complete global model.
6. Conclusion

iSpace using network-based-visions techniques provide promising ways to human-computer interaction through understanding human and object movements from visual data. An important step in achieving this goal is the robust and accurate tracking of the moving objects such as walking human and mobile agents. However, cluttered backgrounds, unknown lighting conditions and multiple moving objects make the tracking tasks challenging. This paper mainly concentrated on global color model based on the local color information for the objects tracking by addressing these difficulties.

Also, this paper presents a representation of global color model based on the proposed Mean Shift Tracking Process (MSTP), in which the structure of the tracking process could be construct with Kalman filter. Then, the local color model and the global color model was proposed based on extracting the objects by background subtraction and creating color histogram. The application of this model achieves the robust tracking of multiple objects seamlessly among different cameras. In the experiments, the comparison method of the global models obtained by different cameras will be described. Effectiveness of the global model for seamless tracking will be also evaluated quantitatively. Integrating discriminant analysis and the Kalman framework, the proposed MSTP algorithm offers a means to relax the assumption of probabilistic structures of data distribution. In addition, the proposed global color-based algorithm is able to tracking a good color space automatically. Some promising color-based tracking results were also achieved by the MSTP approach.

One of the future research directions of the global color model is to explore the nonlinear case of probabilistic methods. In addition, the convergence and stability analysis should be studied in the future work. Currently, the confidence level is an important parameter in the transduction to control the size of labeled set. It needs further studies.

References


**TaeSeok Jin**

He received the B.Sc. degree from Jinju National University, M.Sc. and Ph.D. degrees from Pusan National University, Busan, Korea, in 2000 and 2003, respectively, all in electronics engineering. He is currently a full-time lecturer at the DongSeo University. From 2004 to 2005, he was a Postdoctoral Researcher at the Institute of Industrial Science, The University of Tokyo, Japan. His research interests include multi-sensor fusion, mobile robots, computer vision, and intelligent control. Dr. Jin is a Member of the KFIS, IEEK, ICASE, JSME, and KSME.

Phone : +82-51-320-1541  
Fax    : +82-51-320-1751  
E-mail : jints@dongseo.ac.kr

---

**Hideki Hashimoto (S'83-M'84)**

He received the B.E., M.E., and Dr.Eng. degrees in electrical engineering from The University of Tokyo, Tokyo, Japan, in 1981, 1984, and 1987, respectively. He is currently an Associate Professor at the Institute of Industrial Science, The University of Tokyo. From 1989 to 1990, he was a Visiting Researcher at Massachusetts Institute of Technology, Cambridge. His research interests are control and robotics, in particular, advanced motion control and intelligent control. Dr. Hashimoto is a Member of the Society of Instrument and Control Engineers of Japan, Institute of Electrical Engineers of Japan, and Robotics Society of Japan.

Phone : +81-3-5452-6258  
Fax    : +81-3-5452-6259  
E-mail : hashimoto@iis.u-tokyo.ac.jp