Adaptive Mean Shift Based Face Tracking by Coupled Support Map

Yongwon Hwang1, 2, Mun-Ho Jeong3, Sang-Rok Oh2, and Changyong Yoon4

1 Department of Electrical and Electronic Engineering, Yonsei University, Seoul, Korea
2 Center for Robotics Research, Korea Institute of Science and Technology, Seoul, Korea
3 School of Robotics, Kwangwoon University, Seoul, Korea
4 Department of Electrical Engineering, Suwon Science College, Hwaseong, Korea

Abstract

The mean-shift algorithm is a local search technique that uses the similarity of the color distributed information between the target model and the local candidate image in the target. It has been proven to be superior in simplicity and stability of the technique and has been widely used for face tracking applications. However, one of the major problems in face tracking using color distribution is its vulnerability to backgrounds with similar color distribution, occlusion, and illumination changes. In this paper, we suggest a Coupled Support Map (CSM) to resolve the problem, and show the effectiveness of the robust adaptive mean-shift (AMS) face tracking method. Through series of experiments, we conclude the robustness of suggested algorithms against face size and sudden lighting changes.

Keywords: Mean shift, Face tracking, Occlusion, Depth images, Coupled Support Map

1. Introduction

Many studies have been conducted on effective tracking methods through visual features in complex environments. Among them, objective tracking is one of most challenging problems due to its complexities. Object tracking is to detect the location and size of objects of interest from continuous images and it has many different applications such as monitoring systems [1–4], intelligent user interfaces [5], intelligent sites [6], real-time processing applications such as video compression [7] that requires the ability to track moving objects.

In object tracking problems, the major challenging barrier is to recognize a unique either image or object in various environment to enhance its robustness. Another important aspect is to optimize computational processing time. The mean-shift algorithms have recently attracted attention as a method for real-time tracking of non-rigid bodies based on visual feature values, such as color or texture.

The mean-shift algorithm is a local search technique that uses the similarity of the color distribution between the target model and the local candidate image in the target model and proved to be superior to other methods in its simplicity and stability, and has been widely used in real-time objects.

However, when applied to face tracking using color distribution, there is a disadvantage that it is vulnerable to the influence of object or background having a color distribution similar to the target model. In addition, if the color distribution of the face changes due to the change...
of the illumination or the like, it fails the tracking. Lee et al. [8] has proposed a mean-shift-based face tracking technique that uses distance information as an efficient discrimination between faces and backgrounds and color information as advantages of object discrimination. It shows similar color backgrounds and robustness to occlusion but did not address the methodology for coping with lighting changes.

In this paper, based on the method of [8], we have revised Coupled Support Map (CSM) to enhance the robustness to lighting conditions of tracking algorithm. This is a combination of Depth Support Map (DSM) using distance information, color histogram and radial-distance based Likelihood Support Map. Color histograms using the CSM and bandwidth update in the tracking area enabled robust face tracking even during sudden lighting changes.

This paper briefly discusses the existing mean-shift techniques in the next section, and Section 3 describes the DSM and Likelihood Support Map. Section 4 presents a modified mean-shift scheme based on the Support Map. In Section 5, we show the result of applying to face tracking. Finally we conclude in Section 6.

2. Mean-Shift Tracking

The mean-shift algorithm finds the position that maximizes the similarity of the color histogram between the target model and the candidate model, and is robust to the rotation and shape change of the object. For this reason, it is often used for face tracking.

2.1 Representation of Color Models

The color model is represented using a kernel profile. In order to increase the reliability of the color distribution when the boundary pixels belong to or overlap the background, the kernel profile is a function that assigns a small weight to the pixels located far from the center of the region. When the center position is (0, 0) and the radius of the kernel profile is 1, the target model is expressed as Eq. (1).

\[ q_u = C \sum_{i=1}^{N} k(\|x_i^*\|^2) \delta[b(x_i^*) - u], \quad (1) \]

where \( N \) is the number of pixels belonging to the face region, \( b(x) \) is the kernel function, \( \delta \) is the Kronecker delta function, and \( b(x_i^*) \) is the color value at the normalized position \( x_i^* \) of the pixel \( I \).

Figure 1. 8-Bit color.

The normalization constant \( C \) is determined from the condition \( \sum_{i=1}^{m} q_u = 1 \). In the same way, the color distribution of the candidate region is given by Eq. (2).

\[ \hat{p}_u(y) = C_H \sum_{i=1}^{N_H} k\left((y - x_i)^T \mathbf{H}^{-1}(y - x_i)\right) \delta[b(y) - u] \]

\[ \mathbf{H} = \begin{bmatrix} h_x^2 & 0 \\ 0 & h_y^2 \end{bmatrix}, \quad (2) \]

where \( h_x, h_y \) are the bandwidths of the kernel. The color value \( u \) is an 8-bit value consisting of 3 bits each of red (R) and green (G) and 2 bits of blue (B) as shown in Figure 1.

2.2 Mean-Shift

The object tracking using the general average moving method is expressed using the Epanechnikov kernel, which allocates a smaller weight as the distance from the center \( y \) of the face area increases [9]. Then, the position of the object is estimated by the iterative calculation as in Eq. (3).

\[ y_{t+1} = \frac{\sum_{i=1}^{N_H} x_i w_i g\left((y_t - x_i)^T \mathbf{G}^{-1}(y_t - x_i)\right)}{\sum_{i=1}^{N_H} w_i g\left((y_t - x_i)^T \mathbf{G}^{-1}(y_t - x_i)\right)} \quad (3) \]

Here, \( N \) is the number of pixels belonging to the face region in \( t \) and \( q(x) \) is the differential function of the Epanechnikov kernel. \( w_i \) is a weight of the pixel \( i \), It is expressed by Eq. (4).

\[ w_i = \sum_{k=1}^{m} \sqrt{\frac{\hat{q}_u(y_0)}{\hat{p}_u(y_0)}} \delta[b(x_i) - u], \quad (4) \]

3. Support Map

Conventional mean-shift algorithms, which require a constant color distribution of the tracking object, have limitations in
application in lighting-changing environments. In a dynamic environment where the illumination changes, the similarity between the target model and the candidate model exceeds the allowable range, which increases the possibility of object tracking failure.

It may also cause tracking failures by other backgrounds or objects with color distributions similar to the target model. As a solution to this problem, we have devised a DSM using distance information, a color histogram and a radial-distance-based Likelihood Support Map. The mean-shift algorithm using this is introduced in Section 4.

### 3.1 Depth Support Map

In the case where an object or background similar to the color distribution of the candidate model exists near the candidate model, the possibility of face tracking failure is increased when the face is covered by other objects. In order to solve this problem, the DSM is suggested. The DSM helps to track the face more strongly and clearly when there are objects and backgrounds with similar color distributions in particular.

Assuming that the $i$-th pixel among the pixels constituting the face image in one frame is $x_i$, the DSM $x_i$ corresponding to $s_d(x_i)$ in the current frame is defined as follows:

$$s_d(x_i) = \begin{cases} -1 & \text{if } \bar{d} - d(x_i) < -3\sigma, \\ 1 & \text{if } |\bar{d} - d(x_i)| \leq 3\sigma, \\ 0 & \text{if } \bar{d} - d(x_i) > 3\sigma. \end{cases}$$

$\bar{d}$ is the average of the distance values of the face area determined at time $t - 1$, i.e., the face area at the previous frame, $d_i$ is the distance value at the time $t$, i.e., the $x_i$ position of the candidate model in the current frame, and $D_t$. As a critical point, a value that is usually three times or less the standard deviation of the distance value distribution is used. $D_{face}$ means the pixel $x_i$ exists at a distance similar to the face region, $D_{far}$ is used to distinguish the background from the face, and $D_{close}$ is used to distinguish the object that covers the face.

This DSM is effective when it is used to distinguish faces using distance information, but there is a weakness to obscuring objects having the same distance information. When such an object is occluded, the center position and size of the face region are changed in the face tracking by the conventional mean-shift technique. This can cause problems such as face recognition and face direction estimation performed on the face tracking area. For example, even if the center position and size of the face area are changed by covering the area around the mouth, if the area is regarded as the entire face area, it may lead to errors in face recognition or face direction calculation.

### 3.2 Likelihood Support Map

The tracking area by mean-shift can be expressed as an ellipse area using bandwidth as shown in Eq. (6).

$$d(x_i) = \frac{(x_i - c_x)^2}{h_x^2} + \frac{(y_i - c_y)^2}{h_y^2} \leq 1, \quad (x_i, y_i) \in W_{ex, ey},$$  

where $x_i = (x_i, y_i)$ is the coordinate of pixel $i$ and $(c_x, c_y)$ is the center of tracking area $W_{ex, ey}$. The probability that each pixel belongs to the real face region within this ellipse region becomes larger as it gets closer to the center point, and as the probability of not belonging to the face increases, the probability increases. From this fact, we define the probability based on radial distance as follows:

$$P_r(x_i \mid A = \text{face}) = 1 - d_i(x_i),$$

$$P_r(x_i \mid A = \text{bkg}) = d_i(x_i).$$

Here, $A$ is a random variable representing a region to which a pixel belongs. On the other hand, the probability for each area of the pixel from the color histogram of the face area and the background area obtained from the previous frame is given as follows:

$$P_h(x_i \mid A = \text{face}) = q_{face}(u(x_i)),$$

$$P_h(x_i \mid A = \text{bkg}) = q_{bkg}(u(x_i)).$$

Here, $u(x_i)$ is a color value at $x_i$.

From Eqs. (7)-(9), the likelihood that the pixel belongs to the face area and the background area is

$$P(x_i \mid A = \text{face}) = P_r(x_i \mid A = \text{face})P_h(x_i \mid A = \text{face}),$$

$$P(x_i \mid A = \text{bkg}) = P_r(x_i \mid A = \text{bkg})P_h(x_i \mid A = \text{bkg}),$$

respectively. At this time, the Likelihood Support Map can be defined as follows by the color histogram and the elliptic...
equation.

\[ s_t(x_i) = \begin{cases} 
1 & \text{if } P(x_i \mid A = \text{face}) > P(x_i \mid A = \text{bkg}), \\
0 & \text{otherwise.} 
\end{cases} \]

(13)

The pixel can have a significantly lower probability value in both color histograms of the previous frame, which is due to the sudden illumination change. In order to cope with such illumination changes, the probability value by the color histogram is ignored for stable tracking. In other words,

\[ P_h(x_i \mid A = \text{face}) = P_h(x_i \mid A = \text{face}) = 1, \text{ if } q_{\text{face}}(u(x_i)) < 0.5 \text{ and } q_{\text{bkg}}(u(x_i)) < 0.5. \]

(14)

4. Support Map Based Mean-Shift Algorithm

4.1 Coupled Support Map

Define the CSM as shown in Eq. (15) from the two Support Maps described in the previous chapter. From this, we redefine the pixel weights of Eq. (4) as Eq. (16).

\[ C = \{ c_i \mid c_i = s_d(x_i) \times s_t(x_i), \ i = 1, \ldots, N_H \}, \]

(15)

\[ w'_i = \begin{cases} 
1 & \text{if } c_i = 1, \\
0 & \text{if } c_i = 0, \\
0 & \text{if } c_i = -1. 
\end{cases} \]

(16)

Here, if \( c_i = 1 \), it corresponds to the face area, so the original weight value is used as it is. When \( c_i = 0 \), it corresponds to the face area, but it is occluded, and \( w'_i \) is set to 1 to prevent the position estimation error due to this. If the pixel has a similar color distribution to the target model and belongs to the background, then \( c_i = -1 \). To eliminate this effect in the mean-shift calculation, \( w'_i = 0 \). In the update of the target model described in the next section, only pixels with \( c_i = 1 \) are used. That is, even if it belongs to the face area, it does not consider the hidden pixel.

4.2 Update of Goal Model

As a target model for the mean-shift calculation, the color histogram of the face region must be updated to cope with the illumination change. The problem with adopting this adaptive target model is that it is easy to generate a target model that is distant from the current face region due to error accumulation. This problem can be solved by using the CSM to distinguish pixels in the non-face region shown in Eq. (17). The color histogram of the non-face area in the tracking area is also updated every frame to calculate the Likelihood Support Map shown in Eq. (18).

\[ q_{\text{face}}(u) = C \sum_{i=1}^{N_H} (1 - d(x_i)) \delta[b(x_i) - u] \delta(c_i - 1), \]

(17)

\[ q_{\text{bkg}}(u) = C \sum_{i=1}^{N_H} (1 - d(x_i)) \delta[b(x_i) - u] \cdot \delta(c_i - 1). \]

(18)

Updating bandwidth is essential for updating target models and candidate models [10]. CSM is applied to maximum likelihood based bandwidth calculation method as follows:

\[ H_{t+1} = \frac{\sum_{i=1}^{N_H} \max[\delta(c_i - 1), \delta(c_i)] (x_{i+1} - x_i)(x_{i+1} - x_i)^T}{\sum_{i=1}^{N_H} \max[\delta(c_i - 1), \delta(c_i)]}. \]

(19)

In this case, since \( c_i \) is ‘1’ and ‘0’, it is included in the calculation of the bandwidth since the pixel \( i \) corresponds to the face region.

4.3 Support Map Based Mean-Shift Algorithm

The initialization process of the support map based mean-shift algorithm is based on face detection [11]. That is, from the result of face detection, the tracking area is defined as the face area and the non-face area as shown in Figure 2, and each color histogram is calculated. The follow-up process is as follows:

1. Create Candidate Model in Trace Region
2. Creation of Coupled Support Map
3. Mean-Shift calculation
4. Update color histogram for face area and non-face area (Eqs. (17) and (18))
5. Bandwidth update (Eq. (19))
6. Goto 1
Figure 2. Detection result of face area between red square area and non-face area between blue and red.

Figure 3. Coupled Support Map.

5. Experimental Results

The proposed algorithm is implemented using Intel Core i5 CPU 750 @ 2.67GHz laptop PC and ASUS ‘Xtion PRO LIVE. The real-time input image was 640 × 480 and 30 frames per second. The first row of Figure 3 shows a case in which the distance information is similar but different color distribution from the two cases with similar color but different distance information. The second row shows the Coupled Support Map for each case. White corresponds to 1, gray corresponds to 0, and black corresponds to -1. Through the Coupled Support Map, it can be seen that the pixel belongs to the face area but is divided into the background and the background.

In the color-based face tracking method, the mean-shift tracks the next location through the color distribution between the target and candidate models. Since objects are tracked only through color distribution, there is a risk that an object with a similar color distribution in the tracking area may enter, or that the target may be incorrectly determined for a variety of dynamically changing lighting conditions.

Figure 4 shows that the Coupled Support Map-based mean-shift tracker is robust to illumination changes. The first row in Figure 4 shows the face movements in the rapidly changing illumination, and the second row shows the face tracking results.

Figure 4. Face tracking results in various lighting changes.

The distance between the camera and the face changes, and the bandwidth of the tracker is adjusted appropriately according to the size of the face. The pixel ‘1’ in the Coupled Support Map is white.

6. Conclusion

In this paper, we propose a robust adaptive mean-shift tracking method based on Coupled Support Map when face size changes and illumination change occurs. The existing skin color based face tracking method was difficult to apply in the environment where the skin color model was different from the learning condition, and it was impossible to trace when the illumination changed during the face tracking.

When using the adaptive skin color model, it is possible to cope with general illumination change to some degree, but it has a limit of error accumulation according to time, and it is difficult to cope with excessive illumination change.

The proposed method is similar to the previous studies in terms of adoption of the adaptive color model. However, the face tracking area is divided into the face area and the non-face area, and the robustness is secured by using the Coupled Support Map using the color information and the distance information.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

References

[1] Y. Cui, S. Samarascckera, Q. Huang, and M. Greiffenhagen, “Indoor monitoring via the collaboration between a peripheral sensor and a foveal sensor,” in Proceedings of 1998 IEEE Workshop on Visual Surveillance, Bombay, India, 1998, https://doi.org/10.1109/WVS.1998.646014
[2] J. S. Park, H. T. Kim, and Y. S. Yu, “Video based fire detection algorithm using Gaussian mixture model,” The Journal of the Korea Institute of Electronic Communication Sciences, vol. 6, no. 2, pp. 206-211, 2011.

[3] I. S. Kim and H. S. Shin, “A study on development of intelligent CCTV security system based on BIM,” The Journal of the Korea Institute of Electronic Communication Sciences, vol. 6, no. 5, pp. 789-795, 2011.

[4] I. S. Kim, J. D. Yoo, and B. H. Kim, “A monitoring way and installation of monitoring system using intelligent CCTV under the U-city environment,” The Journal of the Korea Institute of Electronic Communication Sciences, vol. 3, no. 4, pp. 295-303, 2008.

[5] G. R. Bradski, “Real time face and object tracking as a component of a perceptual user interface,” in Proceedings of 4th IEEE Workshop on Applications of Computer Vision, Princeton, NJ, 1998, pp. 214-219. https://doi.org/10.1109/ACV.1998.732882

[6] R. Chellappa, S. K. Zhou, and B. Li, “Bayesian methods for face recognition from video,” in Proceedings of 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing, Orlando, FL, 2002. https://doi.org/10.1109/icassp.2002.1004812

[7] A. Eleftheriadis and A. Jacquin, “Automatic face location detection and tracking for model-assisted coding of video teleconferencing sequences at low bit-rates,” Signal Processing: Image Communication, vol. 7, no. 3, pp. 231-248, 1995. https://doi.org/10.1016/0923-5965(95)00028-U

[8] Y. H. Lee, M. H. Jeong, J. J. Lee, and B. J. You, “Robust face tracking using bilateral filtering,” in Advanced Intelligent Computing Theories and Applications. With Aspects of Theoretical and Methodological Issues, D. S. Huang, D. C. Wunsch, D. S. Levine, and K. H. Jo, Eds. Berlin, Germany: Springer, 2008, pp. 1181-1189. https://doi.org/10.1007/978-3-540-87442-3_146

[9] D. Comaniciu, V. Ramesh, and P. Meer, “Kernel-based object tracking,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 5, pp. 564-577, 2003. https://doi.org/10.1109/TPAMI.2003.1195991

[10] M. H. Jeong, B. J. You, and W. H. Lee, “Color region tracking against brightness changes,” in AI 2006: Advances in Artificial Intelligence, A. Sattar and B. Kang, Eds. Berlin, Germany: Springer, 2006, pp. 536-545. https://doi.org/10.1007/11941439_58

[11] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in Proceedings of 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Kauai, HI, 2001, pp. 511-518. https://doi.org/10.1109/CVPR.2001.990517

[12] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, “Pfinder: real-time tracking of the human body,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 780-785, 1997. https://doi.org/10.1109/34.598236

Yongwon Hwang received the M.S. degree from Department of Electronic Engineering for Control and Biomedical, Konkuk University and he is doing Ph.D. from the Department of Electronic Engineering for Intelligence Control System, Yonsei University, Korea. He has been a researcher at Korea Institute of Science and Technology since 2002. His major interests are robot vision, intelligence, cognitive science, and human-robot interaction.

E-mail: dryongs@gmail.com

Mun-Ho Jeong received the B.S. and M.S. degrees from Korea Advanced Institute and Science and Technology (KAIST), Korea and his Ph.D. from the Department of Mechanical Engineering for Computer Controlled Systems, Osaka University, Japan. He has been a professor at the school of robotics in Kwangwoon University since 2010. His major interests are robot vision and intelligence, image processing, and human-robot interaction.

E-mail: mhjeong@kw.ac.kr

Sang-Rok Oh joined Korea Institute of Science and Technology (KIST) in 1988 after graduation from Korea Advanced Institute of Science and Technology (KAIST) for his Ph.D. degree, and has been working as a principal research engineer at the center for robotics research, KIST. His research interests include information and communication technology,
biorobotics for quality of life, and network based intelligent service robots.
E-mail: sroh@kist.re.kr

Changyong Yoon received B.S., M.S., and Ph.D. degrees in Electrical and Electronic Engineering from Yonsei University, Seoul, Korea, in 1997, 1999, and 2010, respectively. He was a senior research engineer in LG Electronics, Inc., and LG-Nortel, and he developed system software for the DVR and WCDMA from 1999 to 2006. From 2010 to February 2012, he was a chief research engineer in LG Display and developed the circuit and algorithms in touch systems. Since 2012, he has been a Professor in the Department of Electrical Engineering, Suwon Science College, Gyeonggi-do, Korea. His main research interests include intelligent transportation systems, pattern recognition, robot vision, and fuzzy application systems.
E-mail: cyyoon@ssc.ac.kr