Proc. of The 13th Scandinavian Conference on Image Analysis (SCIA2003), (J. Bigun and T. Gustavsson eds.: Image Analysis, LNCS Vol. 2749, Springer), pp. 617–624, 2003.

Tracking Human Heads Based on Interaction between Hypotheses with Certainty

Akihiro Sugimoto[†], Kiyotake Yachi[‡] and Takashi Matsuyama[‡]

[†]National Institute of Informatics, Tokyo 101-8430, Japan
[‡]Graduate School of Informatics, Kyoto University, Japan
sugimoto@nii.ac.jp

Abstract. We propose a method for tracking human heads, where interaction between hypotheses plays a key role. We model appearances of the human head and generate hypotheses for a human head in the image in the model space. We then propagate and reform hypotheses over time in turn to realize tracking human heads. During tracking, we bring about interaction between hypotheses to eliminate the hypotheses denoting false positives and, at the same time, to maintain the hypotheses denoting human heads.

1 Introduction

The ability to detect and track moving people is one of the most important problems in vision. This is because visual interpretation of people and their movements is an important issue in many applications such as visual surveillance and monitoring [2–4, 9].

Condensation [5, 7] was proposed for realizing robust tracking of multi-objects and its effectiveness has been reported [6, 8, 11]. Condensation is the scheme that incorporates stochastic dynamics into the probabilistic framework. In condensation, the density of interpretation samples is normalized to be the probability density. Normalizing the density indicates that interpretation samples in the image are relatively evaluated under the same measure. For the same object, evaluating its interpretation samples under the same measure is effective. For different objects, however, the evaluating measure should be changed depending on the object. This is because the relative evaluation between different objects does not make sense. (If one correct hypothesis with a very high score exists, other correct hypotheses are suppressed. This is irrelevant to them because their correctness should be independent.)

In multi-object tracking, the system does not know the number of objects in the image, and in many cases, the number of objects in the image changes during tracking. The system thus has to identify the correspondence between objects and hypotheses. Once the correspondence is identified, the probabilistic framework such as condensation will work effectively. Without identifying the correspondence, however, we cannot employ the probabilistic framework. To build up a robust and flexible system that is capable of tracking multiple objects even though some of them disappear, new objects come into the image and occlusions occur during tracking, required is a method in which the system maintains

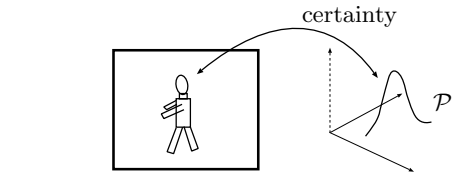


Fig. 1. An object in an image and its hypothesis as the certainty distribution in $\mathcal{P}$.

simultaneous alternative hypotheses of objects and, at the same time, identifies the correspondence between the hypotheses and the objects in the image.

This paper proposes a method for tracking human heads by a monocular camera in which the system maintains simultaneous hypotheses for human heads and establishes the correspondence between the hypotheses and the human heads. To establish such correspondence, we bring about interaction between hypotheses by effectively utilizing both spatial continuity in the image and temporal continuity during tracking.

2 Hypotheses and their representation

2.1 Certainty

The appearance of the human head can be modeled by an ellipse and it has five parameters. Namely, setting the five parameters corresponds to a possible appearance of the human head in the image. Such parameters construct a space, called a *parameter space*, whose dimension is the number of parameters required for the modeling. We denote the parameter space by $\mathcal{P}$.

A human-head appearance in the image corresponds to a point in $\mathcal{P}$. Unless we perfectly and accurately detect a human-head appearance in the image, we have ambiguity in identifying the point in $\mathcal{P}$ that corresponds to the appearance. We thus represent this ambiguity (more exactly, unambiguity) in terms of certainty, where certainty is defined as follows: (i) the domain is $\mathcal{P}$ and the range is $[0, 1]$, (ii) the ambiguity in denoting a human-head appearance is expressed as the value, and (iii) when a point corresponds to a human-head appearance precisely, its value is 1 while it is 0 when a point corresponds a false positive.

Furthermore, to enhance the robustness in the expression of a hypothesis in $\mathcal{P}$, we introduce a distribution, called a *certainty distribution*, over $\mathcal{P}$. That is, for a detected possible human-head appearance (which may include false positives), we generate a hypothesis in the parameter space, where the hypothesis is represented as its certainty distribution (see Fig. 1).

2.2 Certainty-evaluation of an appearance model

For a point in $\mathcal{P}$, we evaluate its corresponding appearance by different features i ($i = 1, 2, \ldots, F$) in the image. They may be color, intensity or gradient information, for example. The evaluation result by feature i is then transformed to

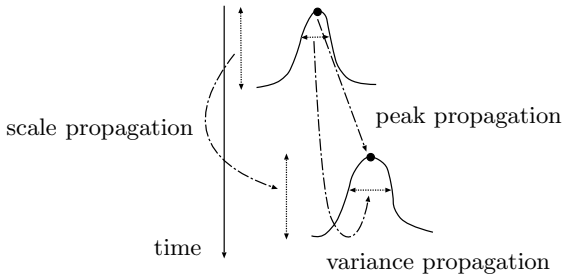


Fig. 2. Propagation of the certainty distribution representing a hypothesis.

certainty. We thus obtain certainty of the point of interest in $\mathcal{P}$. This evaluation process by feature i is called *certainty-evaluation* i. That is, certainty-evaluation i can be expressed by $f^i(\boldsymbol{p}) = \psi^i(e^i(\boldsymbol{p}))$, where $\boldsymbol{p}$ is a point in $\mathcal{P}$, e^i is the evaluation by feature i, and ψ^i is a function that transforms the evaluation result to certainty. Hence, the evaluation of $\boldsymbol{p}$ by all the features is expressed as an integration of f^i's: $f(\boldsymbol{p}) := \kappa \bigotimes_{i=1}^{F} f^i(\boldsymbol{p})$, where $\bigotimes$ implies the integration of f^i's and κ is the normalization factor so that the range of f becomes $[0, 1]$. $\bigotimes$ may be the summation or the multiplication, depending on employed features, in the simplest case.

The transformation function ψ^i can be determined with the help of the ideal value of certainty that is obtained by detecting by hand an appearance of a human head.

2.3 Tracking with certainty distributions

Tracking is conducted in the similar way as condensation. Namely, for a human head, the system generates hypotheses and represents them in $\mathcal{P}$ as certainty distributions and then propagates them over time to predict the hypotheses for the next image. The system reforms the hypotheses through sampling from the newly captured image. The reformed hypotheses are also used for the next propagation. Tracking is realized by the cycle of propagation and reformation of hypotheses. The cycle of propagating and reforming a hypothesis is iterated over time for all hypotheses.

1. **Hypothesis generation**
   To avoid the computational cost for the first detection, we employ the background subtraction. We can, thus, identify the regions within the image that may include objects. Based on the difference of intensities between the input image and the background image, we sample a pixel to obtain position information of an object. This information gives us constraints on $\mathcal{P}$. We randomly sample a point $\boldsymbol{p}$ in $\mathcal{P}$ that satisfies the constraints, and then evaluate $\boldsymbol{p}$ to obtain $f(\boldsymbol{p})$. When $f(\boldsymbol{p})$ is greater than a threshold given in advance, we regard $\boldsymbol{p}$ as a hypothesis and represent it as its certainty distribution using $f(\boldsymbol{p})$. We iterate this sampling to generate hypotheses of objects. We remark

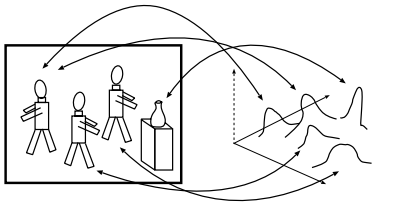


Fig. 3. Relationship between certainty distributions in parameter space and objects in the image.

that different hypotheses may be generated for one object appearance and that a hypothesis does not necessarily denote an object in the image due to false positives.

2. **Hypothesis propagation**
   For a hypothesis of an object, its certainty distribution is not static over time. In accordance with the object movement, it dynamically changes. We incorporate the object movement into the propagation of a certainty distribution. Namely, we propagate the certainty distribution for the current image to obtain the certainty distribution for a new image. The propagation over time is conducted by the parameters representing the certainty distribution. For example, they are the peak, the variance and the scale of the distribution as shown in (Fig. 2).
3. **Hypothesis reforming**
   A propagated hypothesis does not reflect the information within a newly captured image. To accurately reflect the information within the new image, the propagated hypothesis has to be reformed through feedback from the new image. The propagated hypothesis allows us to know certainty for each point in $\mathcal{P}$. We therefore sample points based on this certainty to have a sample set for the hypothesis. Then, in the new image we evaluate the sample points to reform the hypothesis.

3 Interaction between hypotheses

3.1 Hypotheses and their denoting objects

When tracking human heads, the number of people in the image is unknown. Moreover, the number may change over time. Namely, some objects may disappear from the image and some may newly come into the image. To robustly track human heads under these conditions, we have to identify the relationship between hypotheses and their corresponding human heads in the image (Fig. 3). We also have to identify false positives.

For the discussion below, we classify the hypotheses maintained over time into three types. We remark that the system itself has not identified the type to which a hypothesis belongs.

Type A: the hypotheses denoting a human head (human head A).
Type $\bar{\mathbf{A}}$: the hypotheses denoting other human heads.
Type F: the hypotheses denoting false positives.

3.2 Identifying false positives through cumulation of certainty

To identify hypotheses denoting false positives, we use temporal continuity of certainty during tracking.

For a hypothesis h at time t_k, we compute the maximum value of its certainty $\gamma_{t_k}^h$. If the hypothesis h denotes a human head, namely, h does not belong to type F, cumulation of $\gamma_{t_k}^h$ over t_k increases at a high speed and it does not become saturated as far as the human head exists in the image[1]. If h is a false positive, on the other hand, cumulation of $\gamma_{t_k}^h$ slowly increases and can be saturated.

We set a threshold $\tilde{\gamma}$ that expresses the likeness of a human head, and cumulate over time the difference of $\gamma_{t_k}^h$ from $\tilde{\gamma}$. If the cumulation of the difference over a fixed amount of time is not greater than another threshold $\tilde{\Gamma}$ for the cumulation, we then regard the hypothesis as a false positive and throw it away.

3.3 Identifying objects through interaction between hypotheses

When the projections of two hypotheses not in type F onto the image become sufficiently close to each other, two cases can occur in the image depending on their denoting human heads: one is the case where the human heads are the same, and the other is the case where they are different. In the former case, the two hypotheses should be merged since we now know that the two hypotheses are for the same human head. In the latter case, on the other hand, two different human heads have become sufficiently close to each other in the image. This implies that one human head is going to occlude the other. Hence, both the two hypotheses should survive. To properly deal with these situations, the system brings about interaction between hypotheses close to each other where spatial continuity in the image plays an important role.

We incorporate fringe information of a hypothesis to measure interaction between two hypotheses. Here, fringe information of a hypothesis implies features, called *fringe features*, obtained nearby the appearance in the image that corresponds to the peak of the certainty distribution expressing the hypothesis. Fringe features can be detected, for example, from the neck line, the shoulder line and even from the body.

We introduce a measure for the degree of interaction between two hypotheses based on their fringe features. For a fringe feature, we frame-wisely compute the distance between the fringe features of the hypotheses. A small distance of a fringe feature supports the sameness of two human heads and, therefore, the two hypotheses should be merged. A large distance, on the other hand, supports

[1] In the case where a human head disappears from the image, it is expected that cumulation of $\gamma_{t_k}^h$ starts being saturated synchronized with time when the human head disappears. h then becomes a false positive.

Table 1. The desired results after the interaction between two hypotheses h and h'.

type of h	A	A	A	F
type of h'	$\bar{\text{A}}$	A	F	F
surviving type	A and $\bar{\text{A}}$	A	A	F
sign of w	+ (large)	−	± (small)	−

the difference between the two human heads and, thus, both the two hypotheses should survive. The distance itself can be regarded as the degree of interaction for the frame of interest. A weighted average of the degrees of interaction over the frames captured so far, gives us the degree of interaction measured by the fringe feature. We thereby integrate the degree of interaction measured by each fringe feature to obtain the degree of interaction between the two hypotheses at that time.

To realize the above measure, we define w such that w is negative if two hypotheses should be merged, and w is positive if both should survive. When w is equal to zero, two hypotheses do not interact. Table 1 enumerates all the cases where two hypotheses interact with each other ($\pm$ implies that w can be either positive or negative depending on a hypothesis in type F.)

The cumulative certainty of a hypothesis h is evaluated by cumulating over time, certainty of h itself and the degree of interaction from other hypotheses that interact with h. A hypothesis with a positive w contributes to increase cumulative certainty of h, and a hypothesis with a negative w contributes to decrease cumulative certainty of h. A hypothesis that makes w zero, contributes nothing to cumulative certainty of h (no interaction occurs).

This mechanism of interaction between hypotheses leads to merging and surviving of hypotheses that we expect. The system eventually identifies the one-to-one correspondence between hypotheses and human heads.

4 Experiments

We employed three features [1, 10] to evaluate the human-head model, i.e., the ellipse: (i) the intersecting ratio between the chromaticity-based color histogram inside the ellipse and that of the human head, (ii) the normalized mean of the vertical components of the gradient magnitude around the perimeter of the ellipse, (iii) the average of intensity difference inside the ellipse from the background.

We brought about interaction between hypotheses when the foots of their certainty distributions intersected, and used two fringe features: the velocity in the image of the ellipse center and the histogram of gradient magnitude inside the rectangle constructed below the ellipse. Here, gradients were computed along the vertical direction in the image.

We generated the following situation. At first, one person A comes in the image (#34) and after passing by the front of the camera, A disappears (#59). Next, another person B comes in (#97) and stops with captured around in the

Fig. 4. Images (with frame numbers) in the outdoor and detected human heads.

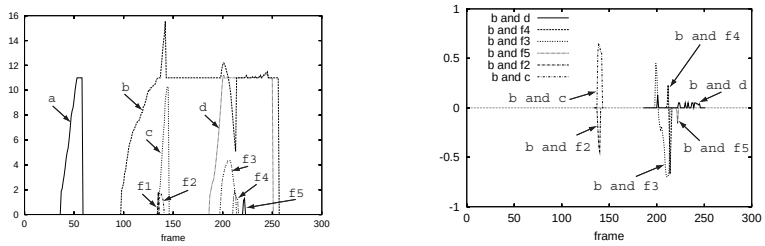
center of image (#110). Another person C then comes in (#134) and passes by behind B (#140) and then disappears (#146). Next, a fourth person D comes in (#183) and becomes very close to B but does not pass by (#224). After B and D part from each other (#243), they disappear one after another (#251 and #257). (Here, the frame numbers are attached.) We remark that the velocity of their movements is just like walking naturally.

Figure 4 shows an example of acquired image sequence. The ellipses corresponding to the peaks of certainty distributions of generated hypotheses in each frame are superimposed on the frames as detected human heads. Labels of the ellipses represent the names of generated hypotheses during tracking. Fig. 4 indicates that ellipses b, c and d correctly denote the head of the persons B, C, and D, respectively. We see that B, C and D are almost correctly tracked even under occlusions and the change in numbers of human heads.

Cumulative certainty of generated hypotheses is shown in Fig. 5 (a). We see that nine hypotheses were generated and that five of them were identified to be false positives. The degree of interaction, i.e., w, between hypothesis b and the other hypotheses is shown in Fig. 5 (b). Here, $f2$, $f3$, $f4$ and $f5$ are false positives, all of which were generated near b. We see that interaction between hypotheses are brought about as we expected and that the system correctly maintain hypotheses.

5 Concluding remarks

To track multiple objects, the system has to maintain simultaneous alternative hypotheses. Without knowing the correspondence between hypotheses and objects, hypotheses should be evaluated not relatively but absolutely. Certainty introduced in this paper is motivated by this observation.



(a) cumulative certainty (b) degree of interaction with hypothesis b

Fig. 5. Cumulative certainty and the degree of interaction between hypotheses.

To establish the correspondence between hypotheses and objects, interaction between the hypotheses is indispensable. This is because we have to eliminate the hypotheses denoting false positives and, at the same time, to maintain the hypotheses denoting the objects even if the number of object changes and occlusions occur.

Acknowledgements. This work is in part supported by Grant-in-Aid for Scientific Research of the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

1. S. Birchfield: Elliptical Head Tracking Using Intensity Gradients and Color Histograms, *Proc. of CVPR '98*, pp. 232–237, 1998.
2. Y. Cui, S. Samarasekera, Q. Huang and M. Greiffenhagen: Indoor Monitoring via the Collaboration between a Peripheral Sensor and a Foveal Sensor, *Proc. of the IEEE Workshop on Visual Surveillance*, pp. 2–9, 1998.
3. L. Davis, S. Fejes, D. Harwood, Y. Yacoob, I. Hariatoglu and M. J. Black: Visual Surveillance of Human Activity, *Proc. of the 3rd ACCV*, Vol.2, pp. 267–274, 1998.
4. D.M. Gavrila: The Visual Analysis of Human Movement: A Survey, *Computer Vision and Image Understanding*, 73 (1999), 1, pp 82–98.
5. M. Isard and A. Blake: Contour Tracking by Stochastic Propagation of Conditional Density, *Proc. of the 4th ECCV*, Vol. 1, pp. 343–356, 1996.
6. M. Isard and A. Blake: ICondensation: Unifying Low-Level and High-Level Tracking in s Stochastic Framework, *Proc. of the 5th ECCV*, Vol. 1, pp. 893–908, 1998.
7. M. Isard and A. Blake: Condensation–Conditional Density Propagation for Visual Tracking, *Int. J. of Computer Vision*, 29 (1998), 1, pp. 5–28.
8. J. MacCormick and A. Blake: A Probabilistic Exclusion Principle for Tracking Multiple Objects, *Proc. of the 7th ICCV*, pp. 572–578, 1999.
9. T. Matsuyama: Cooperative Distributed Vision —Dynamic Integration of Visual Perception, Action, and Communication —, *Proc. of IUW*, pp. 365–384, 1998.
10. K. Yachi, T. Wada and T. Matsuyama: Human Head Tracking using Adaptive Appearance Models with a Fixed-Viewpoint Pan-Tilt-Zoom Camera *Proc. of the 4th FG*, pp. 150-155, 2000.
11. S. Zhou, V. Krueger and R. Chellappa: Face Recognition from Video: A CONDENSATION Approach, *Proc. of the 5th FG*, pp. 221-226, 2002.