Information processing apparatus and method for recognizing a user's emotion

US 8407055 B2

摘要

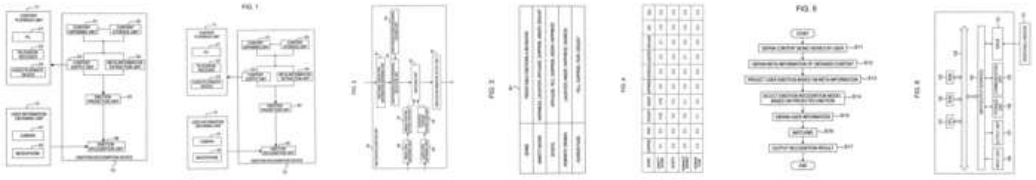
An information processing apparatus includes an obtaining unit that obtains meta-information concerning content; a predicting unit that predicts an emotion of a user who is viewing the content from the meta-information obtained by the obtaining unit; and a recognizing unit that recognizes an emotion of the user using the emotion predicted by the predicting unit and user information acquired from the user.

公开号	US8407055 B2
发布类型	授权
专利申请号	US 11/498,348
公开日	2013年3月26日
申请日期	2006年8月3日
优先权日	2005年8月5日
缴费状态	已支付
公告号	CN1908965A, CN100394438C, US20070033050
发明者	Yasuharu Asano, Noriyuki Yamamoto
原受让人	Sony Corporation
导出引文	BiBTeX, EndNote, RefMan

专利引用 (31), 非专利引用 (3), 被以下专利引用 (36), 分类 (10), 法律事件 (2)

外部链接: 美国专利商标局 (USPTO), 美国专利商标局 (USPTO) 专利转让信息, 欧洲专利数据库 (Espacenet)

图片 (7)



说明

CROSS REFERENCES TO RELATED APPLICATIONS

The present invention contains subject matter related to Japanese Patent Application JP 2005-227527 filed in the Japanese Patent Office on Aug. 5, 2005, the entire contents of which are incorporated herein by reference.

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to information processing apparatuses and methods, and programs. More specifically, the present invention relates to an information processing apparatus and method for recognizing a user's emotion based on information attached to content and information such as facial expressions of the user, and to a program used therewith.

2. Description of the Related Art

Techniques for recognizing (or estimating) a person's (or user's) emotion based on voice tone or facial expressions of the user have been proposed (see, for example, Japanese Unexamined Patent Application Publication No. 10-228295 and MATSUMOTO, et. al., "Emotion Recognition Using Face Image and Speech Information for Robots", The 22nd Annual Conference of Robotics Society of Japan).

SUMMARY OF THE INVENTION

In the methods of the related art for recognizing a user's emotion, voice of the user is obtained by using a microphone, and the emotion of the user is recognized based on the tone or the like of the voice. Further, an image of the face of the user is captured by using a camera, and the emotion of the user is recognized based on a facial expression obtained from the captured image.

In the approach for recognizing a user's emotion based on speech and image

权利要求 (13)

What is claimed is:

1. An information processing apparatus for recognizing a user's emotion, said apparatus comprising:

obtaining means for obtaining meta-information concerning content;

predicting means for predicting an emotion of a user who is viewing the content from the meta-information obtained by the obtaining means; and

recognizing means for recognizing an emotion of the user using the emotion predicted by the predicting means and user information acquired from the user, wherein said recognizing means comprises a prediction information obtaining means, a recognition-model selection means, a recognition-model holding means, and a matching means, wherein a method for recognizing a user's emotion is based on the Bayes decision rule,

wherein said prediction information obtaining means obtains prediction information from said predicting means and supplies said prediction information to said recognition-model selection means to select appropriate recognition model(s) from said recognition-model holding means based on said prediction information, and supplies the selected recognition model(s) to said matching means,

wherein an emotion recognition result output from said emotion recognizing means is fed back to said emotion predicting means so that said emotion predicting means can be adapted to said user, and wherein the emotion recognition result is used as information indicating the user's preference to obtain information suited to the preference of the user when obtaining information via a network, wherein the emotion is an emotion of a user who is viewing the content.

erroneously determined (that is, the recognition accuracy is low). Thus, it is desirable to recognize the emotion of the user with higher accuracy (that is, it is desirable to more reliably estimate the emotion of the user).

It is therefore desirable to achieve high-accuracy recognition (or estimation) of a user's emotion.

An information processing apparatus according to an embodiment of the present invention includes obtaining means for obtaining meta-information concerning content; predicting means for predicting an emotion of a user who is viewing the content from the meta-information obtained by the obtaining means; and recognizing means for recognizing an emotion of the user using the emotion predicted by the predicting means and user information acquired from the user.

The recognizing means may perform weighting for the emotion of the user according to the emotion predicted by the predicting means.

The information processing apparatus may further include sound-feature-value extracting means for collecting sound emitted from the user and extracting a feature value from the collected sound. The recognizing means may use the feature value extracted by the sound-feature-value extracting means as the user information.

The information processing apparatus may further include image-feature-value extracting means for capturing an image of the user and extracting a feature value from the captured image. The recognizing means may use the feature value extracted by the image-feature-value extracting means as the user information.

The meta-information obtained by the obtaining means may be text information concerning the content.

When the content includes audio information, the obtaining means may extract a feature value from the audio information, and may obtain the extracted feature value as the meta-information.

When the content includes speech information, the obtaining means may recognize speech based on the speech information, and may obtain text extracted from the recognized speech as the meta-information.

When the content includes video information, the obtaining means may obtain information concerning performers appearing in video based on the video information as the meta-information.

When the content has time information attached thereto, the obtaining means may extract the meta-information in accordance with the progress of the content according to the time information.

The predicting means may include a table used to predict the emotion of the user. The emotion of the user recognized by the recognizing means may be supplied to the predicting means, and the predicting means may update the table in response to the supplied recognized emotion of the user.

An information processing method according to another embodiment of the present invention includes the steps of obtaining meta-information concerning content; predicting an emotion of a user who is viewing the content from the obtained meta-information; and recognizing an emotion of the user using the predicted emotion and information acquired from the user.

A program according to another embodiment of the present invention causes a computer to execute a process including the steps of obtaining meta-information concerning content; predicting an emotion of a user who is viewing the content from the obtained meta-information; and recognizing an emotion of the user using the predicted emotion and information acquired from the user.

In an embodiment of the present invention, a user's emotion predicted from content being viewed by the user and information acquired from the user who is viewing the content are used to recognize the emotion of the user.

According to an embodiment of the present invention, therefore, a user's emotion can be recognized

2. The information processing apparatus according to claim 1, wherein the recognizing means performs weighting for the emotion of the user according to the emotion predicted by the predicting means.

3. The information processing apparatus according to claim 1, further comprising

sound-feature-value extracting means for collecting sound emitted from the user and extracting a feature value from the collected sound, wherein the recognizing means uses the feature value extracted by the sound-feature-value extracting means as the user information.

4. The information processing apparatus according to claim 1, further comprising

image-feature-value extracting means for capturing an image of the user and extracting a feature value from the captured image, wherein the recognizing means uses the feature value extracted by the image-feature-value extracting means as the user information.

5. The information processing apparatus according to claim 1, wherein the meta-information obtained by the obtaining means comprises text information concerning the content.

6. The information processing apparatus according to claim 1, wherein when the content includes audio information, the obtaining means extracts a feature value from the audio information, and obtains the extracted feature value as the meta-information.

7. The information processing apparatus according to claim 1, wherein when the content includes speech information, the obtaining means recognizes speech based on the speech information, and obtains text extracted from the recognized speech as the meta-information.

8. The information processing apparatus according to claim 1, wherein when the content includes video information, the obtaining means obtains information concerning performers appearing in video based on the video information as the meta-information.

9. The information processing apparatus according to claim 1, wherein when the content has time information attached thereto, the obtaining means extracts the meta-information in accordance with the progress of the content according to the time information.

10. The information processing apparatus according to claim 1, wherein:

the predicting means includes a table used to predict the emotion of the user;

the emotion of the user recognized by the recognizing means is supplied to the predicting means; and

the predicting means updates the table in response to the supplied recognized emotion of the user.

11. An information processing method carried out on an information processing apparatus for recognizing a user's emotion, the method comprising the steps of:

obtaining, via a processor, meta-information concerning content from an obtaining unit; predicting an emotion of a user who is viewing the content from the obtained meta-information using a predicting unit; and

recognizing using the processor, an emotion of the user using the predicted emotion and information acquired from the user by a recognizing unit, wherein said recognizing step further comprises the steps of:

obtaining prediction information from said predicting step;

supplying said prediction information to a recognition-model selection step to select appropriate recognition model(s) from a recognition-model holding