A Study on Gaze Estimation Using Head and Body Pose Information

Nobuhiro Funatsu[1], Tomokazu Takahashi[2], Daisuke Deguchi[3], Ichiro Ide[1], and Hiroshi Murase[1]
[1] Graduate School of Information Science, Nagoya University, Furou-cho, Chikusa-ku, Nagoya-shi, Aichi-ken, Japan
Email: funatsun@murase.m.is.nagoya-u.ac.jp, {ide, murase}@is.nagoya-u.ac.jp
[2] Faculty of Economics and Information, Gifu Shotoku Gakuen University, 1-38, Nakauzura, Gifu-shi, Gifu-ken, Japan
Em ail: ttakahashi@gifu.shotoku.ac.jp
[3] Information and Communications Headquarters, Nagoya University, Furou-cho, Chikusa-ku, Nagoya-shi, Aichi-ken, Japan
Email: ddeguchi@nagoya-u.jp

***Abstract*—Gaze estimation from an image is an important technique for tasks such as driver monitoring and measuring of advertising effectiveness. For this, most existing methods require a high quality image of eyes. However, It is difficult to obtain the eye images when eye occlusion occurs due to sunglasses or face rotation. Another approach approximates head directions to gaze directions. However, the gaze direction is affected not only by the head direction but also by the body direction. Thus, we are studying a method that estimates the gaze direction accurately using information on both head and body pose directions. Experimental result showed that the proposed method could estimate gaze directions more accurately than by using the information on only head directions.**

***Index Terms*—gaze estimation, head direction, body direction**

I. Introduction

Gaze estimation techniques using a camera have widely been studied in recent years. The techniques can be applied to tasks such as driver monitoring for a driving support system or measuring of advertising effectiveness. In a driving support system, the gaze measurements are used for warning on ahead in order to prevent a traffic accident. For example, the system warns a driver before he/she enters an intersection ignoring a traffic signal, or makes a lane change or a turn without checking the mirror. The gaze measurements are also considered useful from the viewpoint of filtering information provided to a driver. Traffic signal, pedestrian and vehicle ahead detection using an in-vehicle camera is actively studied. However, increasing the amount of information the driver receives from the systems distract, the driver and may even lead to the increase of traffic accidents. In order to avoid distraction, the driving support system should warn only when the driver is not focusing to important targets. On the other hand, techniques to measure the effectiveness of outdoor advertisements based on gaze measurements have also been studied. The amount of traffic in front of an advertisement is traditionally used to measure the advertising effectiveness. However, it does not necessarily correspond to the number of people who have actually recognized the advertisement. For this reason, there is a need for a technique to measure it accurately. Moreover, in recent years, advertising media using a flat panel monitors, called digital signages, is rapidly becoming popular. Since digital signages can change its contents from time to time, real-time measurement of the advertising effectiveness is also required. To achieve this, a technique [1] using a camera mounted near the advertisement is developed. This technique counts the number of persons who see the advertisement and the time while the person focuses on it by detecting the face and estimating its direction.

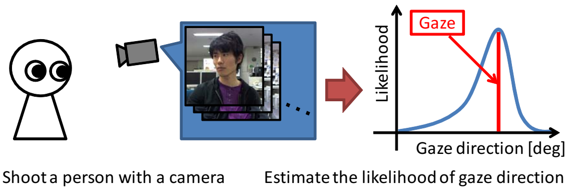

Gaze
Likelihood
Gaze direction [deg]
Shoot a person with a camera
Estimate the likelihood of gaze direction

Fig. 1. Gaze estimation by images.

Most existing methods to estimate a gaze direction using a camera uses the location of the iris in the eye area. The methods require a high-quality image of the eye captured from a high resolution camera or a wearable device. However, an eye image obtained under real environments is often in low resolution, affected by illumination changes and occluded by sunglass or head rotation. On the other hand, since head pose estimation is easier than the detection of eyes, another approach is used in [2] that approximates a head direction to a gaze direction. This approach assumes that the gaze direction is probabilistically distributed around the face direction. However, it has the following two problems:

1) Factors that influence a gaze direction are not only the face direction but also the body direction, their motion, and the situation the person is in. For example, since the speed and the motion range differ between eye and neck, head pose with respect to body pose and these motion bias the shape of the distribution of gaze directions.
2) Results of pose estimation from an image always obtain unavoidable errors due to low resolution, illumination, occlusion, and so on.

In this paper, aiming to achieve accurate gaze estimation in a situation where eyes could not be detected from an image accurately, we report a study on a gaze estimation method using a camera with the following approaches:

1) We particularly focus on the head pose and the body pose for the gaze estimation, and use the relationship between the gaze direction and the head pose with respect to the body pose that is obtained from the result of a human experiment.
2) We use a likelihood distribution of pose information for the gaze estimation instead of directly using a pose estimation result.

The rest of the paper is organized as follows: Section 2 gives a brief overview of related work. The proposed method is described in section 3 and the experimental results are discussed in section 4. Finally section 5 gives conclusions of this paper and future work.

II. RELATED WORK

There are methods to estimate the gaze direction from an image of a person by detecting eyes. The methods are roughly divided into two; methods based on a three dimensional eye model and those based on the appearance of eyes. To estimate the gaze direction, the methods based on the three dimensional eye model estimate the position of the iris by detecting its edges and fitting an ellipse to it, and then fitting the estimated position to the eye model [3], [4]. On the other hand, the methods based on the appearance do not detect the iris, but collects many learning images of the eye in various gaze directions and learns the distribution of the pixel values to estimate the gaze direction. As the methods based on the appearance, methods using neural network [5], [6] and a method using nearest neighbor search [7] are proposed. These methods require high quality images of eyes. Accordingly, the person has to approach the camera or wear a special device. However, the use of wearable devices is not practical especially when considering the measurement of advertising effectiveness. Even if the person approaches the camera, obtaining high-quality eye images is difficult due to illumination changes and occlusion.

As an approach to estimate the gaze without observing the person, a method using a saliency map [8] is proposed. The saliency map is a calculation model of a visual attention based on study results [9] in the field of cognitive science. In the method, a position that a person is likely to notice is estimated from image features computed from an input image. In order to estimate the gaze direction, a first-person-viewpoint video is used in [10]. The video is captured by the camera fixed on the head to compute the saliency map. The position that a driver is likely to notice is estimated in [2] by combining the saliency map from a video captured with a panorama camera attached to a car and the head pose from a video capturing the driver.

III. PROPOSED METHOD

The head pose, the body pose, their motion, and so on. are considered as factors that influence the gaze direction. Among them, we focus in this paper the head pose and the body pose. Besides, we only consider the horizontal gaze direction in this paper.

Even if we can obtain accurate pose information, we cannot determine the gaze direction uniquely because eyes can move independently. In order to represent this ambiguity, we represent the gaze direction under given pose information as a conditional likelihood function $L(\text{gaze}|\text{pose information})$. It is difficult to estimate accurate pose information from an image of a person because the image is often in low resolution, affected by illumination changes, and partial occlusions. Therefore, we represent pose information when the image of a person is obtained as a conditional likelihood function $L(\text{pose information}|\text{image of person})$. These two conditional likelihood functions are integrated to obtain the final gaze estimation result. Figure 2 shows the process flow of the proposed method. The proposed method mainly consists of three parts that are an online gaze estimation part and two offline learning parts. Each part is explained below.

A. Gaze estimation

We represent a gaze direction as a scalar x, pose information as a two dimensional vector $\boldsymbol{y}$ consisting of head and body directions, and an observed image of a person as a vector $\boldsymbol{z}$. In order to estimate x from the images of person $\boldsymbol{z}$, the proposed method calculates the following form:

$$L(x|\boldsymbol{z}) = \sum_{\boldsymbol{y}} L(x|\boldsymbol{y})L(\boldsymbol{y}|\boldsymbol{z}). \tag{1}$$

The gaze direction that maximizes the likelihood value is finally output as a result of gaze estimation.

In the online gaze estimation stage, the proposed method first calculates the likelihood distribution of the pose information $L(\boldsymbol{y}|\boldsymbol{z}')$ from the input image $\boldsymbol{z}'$ by the pose estimator constructed in an offline process. Integrating the resulting likelihood $L(\boldsymbol{y}|\boldsymbol{z}')$ and the relationships between gaze and pose information $L(x|\boldsymbol{y})$ acquired in an offline process, the proposed method obtains the likelihood distribution of the gaze $L(x|\boldsymbol{z}')$ in which the ambiguity of pose estimation and the relationship between the gaze and pose information are considered.

B. Construction of pose estimator

Likelihood functions of the pose information conditioned by the captured image, $L(\boldsymbol{y}|\boldsymbol{z})$, are calculated by an existing pose estimation method. In the paper, the pose information $\boldsymbol{y}$ consists of the face direction and the body direction, and the body direction is fixed to 0 degree, therefore, only the face direction is estimated based on an eigenspace method. This method first calculates an eigenspace from learning face images and projects all of the learning images on the eigenspace. To estimate the face direction, an input image is also projected on the eigensapce, and distances between the input image and the learning images for each direction $d(\boldsymbol{y}|\boldsymbol{z}')$ are calculated. To obtain the likelihood values, the distances for all face directions are then transformed by the following

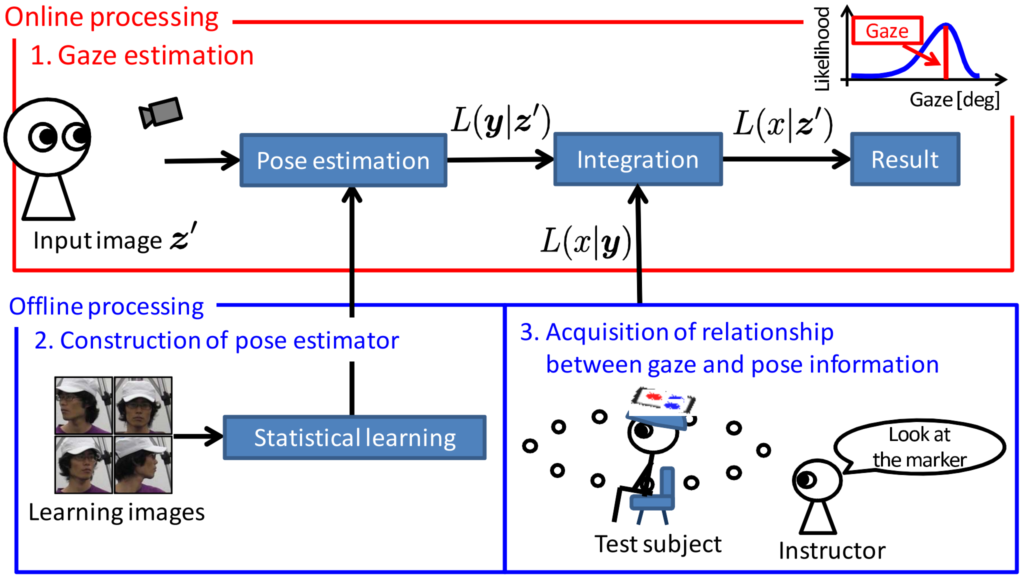


Fig. 2. Process flow of the proposed method.

formula:

$$L(\boldsymbol{y}|\boldsymbol{z}') = \frac{S(\boldsymbol{y}|\boldsymbol{z}')}{\sum_{\boldsymbol{y}} S(\boldsymbol{y}|\boldsymbol{z}')}, \tag{2}$$

$$S(\boldsymbol{y}|\boldsymbol{z}') = \max_{\boldsymbol{y}} d(\boldsymbol{y}|\boldsymbol{z}') - d(\boldsymbol{y}|\boldsymbol{z}'). \tag{3}$$

We regard the transformed values as the likelihood distribution $L(\boldsymbol{y}|\boldsymbol{z}')$.

C. Acquisition of relationships between the gaze and the head pose with respect to the body pose

We acquired the relationships between the gaze direction and the head pose with respect to the body pose by analyzing sets of gaze, head and body directions that were obtained by the following experiment.

1) Experimental overview: The overview of the experiment is shown in Figure 3. First, we arranged 36 markers in 10 degrees horizontal angular interval along a circle with a radius of 2 meters, and put a chair at the its center. To fix a body direction in 0 degree, a test subject sat down in the chair. To measure a head direction, a color marker was attached to the head of the test subject. We obtained the head direction by detecting the color marker by an above camera when the test subject looked at the indicated marker. Ten subjects participated to the experiment, and each subject were asked to look at the 36 markers 3 times.

2) Experimental result: After quantizing the head poses with respect to the body pose obtained by the experiment to 36 directions, we calculated the means and standard deviations of the gaze directions for each quantized head pose. The result is shown in Figure 4. In this figure, the horizontal axis indicates the gaze direction, and the vertical axis indicates the head pose with respect to the body pose. We plotted means of gaze by rhomboid points and standard deviations by line segments. For comparison, the gaze direction when the head direction was assumed as the gaze direction was plotted by a red solid line. Figure 5 shows distributions of the gaze direction when face directions with respect to the body directions were 0, 40, 80 and 120 degrees. In these figures, red dash lines represent the face directions. From Figures 4 and 5, when the face direction with respect to the body direction is 0 degree (the face and the body are in the same direction), we can see that the gaze direction distributes around the face direction. This trend agrees with the hypothesis by Doshi and Trivedi [2] that the gaze is distributed around the face direction (rhomboid points are plotted on the red line). In contrast, we also confirmed that the gap between the gaze and the face direction widened and the variance became small when angles between head and body directions widened. This indicates that the use of the body direction in addition to the head direction would work effectively to achieve accurate gaze estimation.

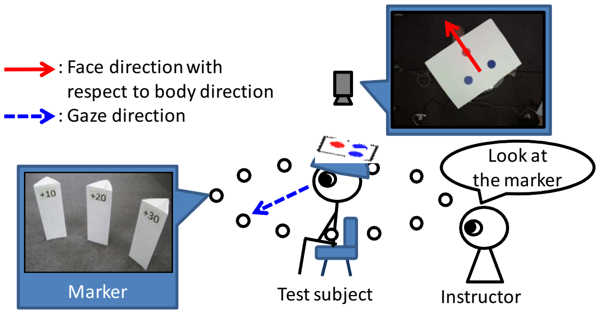
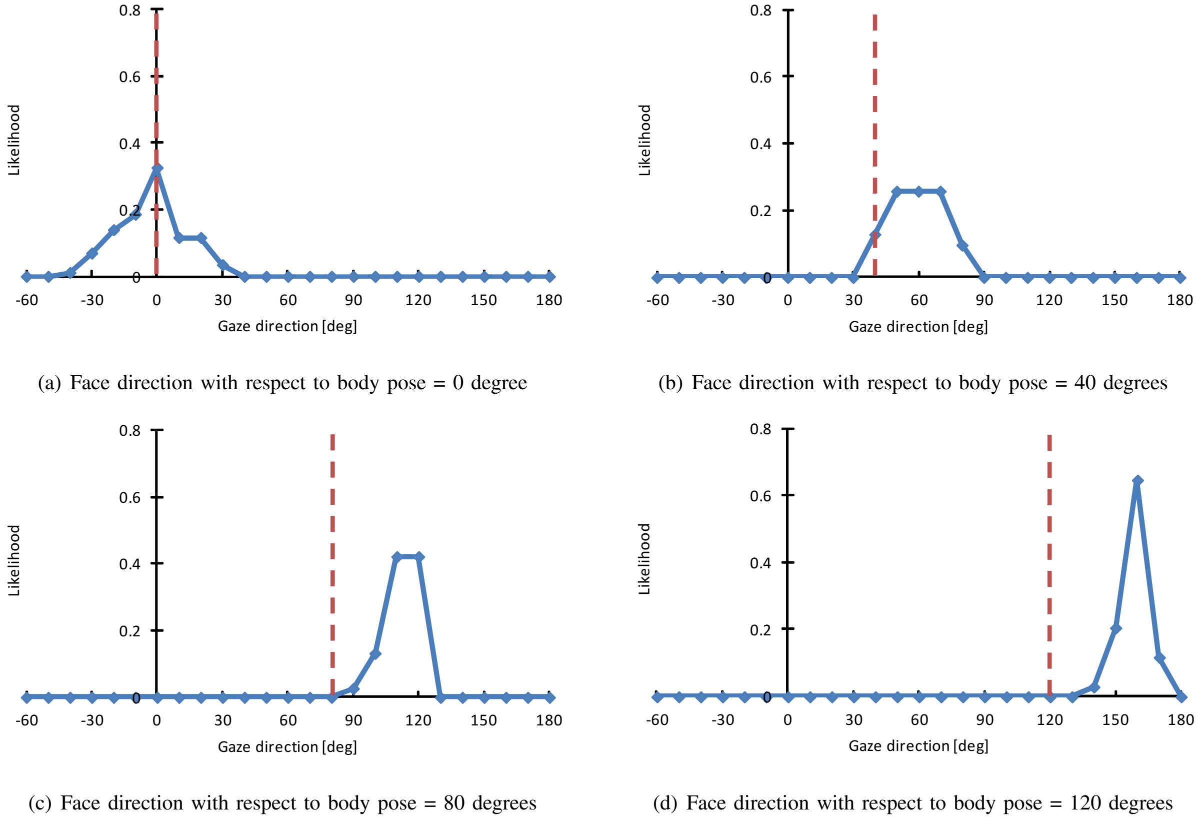


Fig. 3. Overview of the experiment.

IV. Experiment

In order to investigate the effectiveness of the proposed method, we conducted an experiment of the gaze estimation. To evaluate the gaze estimation accuracy, we measured



(a) Face direction with respect to body pose = 0 degree

(b) Face direction with respect to body pose = 40 degrees

(c) Face direction with respect to body pose = 80 degrees

(d) Face direction with respect to body pose = 120 degrees

Fig. 5. Part of the likelihood function of the gaze conditioned by the face direction with respect to body direction.

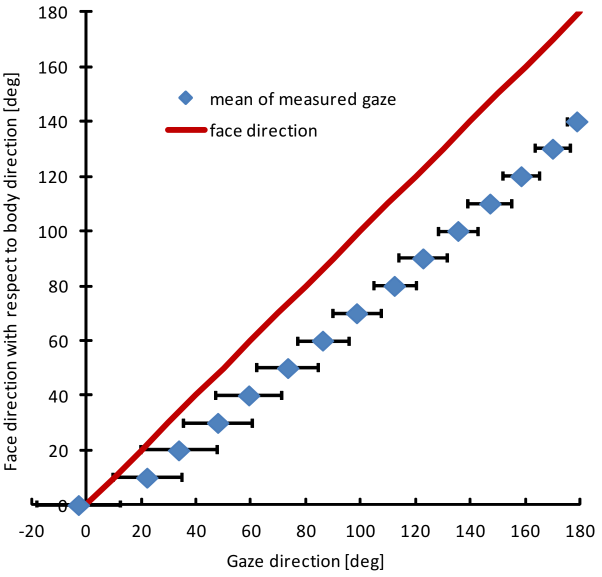


Fig. 4. Relationship between the gaze direction and the face direction with respect to body direction.

the Root Mean Square Error (RMSE) between estimations by the proposed method and the acted measurements. For comparison, we also evaluated the accuracy of an existing method which approximates the estimated head direction to the gaze. In the experiment, we used likelihood functions $L(x|\boldsymbol{y})$ obtained from the experiment described in section 4. In order to obtain learning images of the face pose estimator, a subject was captured by a camera in front of the subject and face areas in the images were cropped manually. Similarly, we obtained test images and cropped face areas of the images manually. The test images were composed of 108 images (= 36 directions × 3 images). Figure 6 shows the estimation error for each gaze direction. From the figure, we can see that the accuracy of the proposed method was higher than that of the existing method in most cases. Thus, we confirmed the effectiveness of the use of the body pose in addition to the face pose for gaze estimation. However, when the gaze directions were around 0 degree, the estimation errors of the proposed method were larger than those of the existing method. This is because the proposed method output a larger value as the result of the gaze estimation when the pose estimator failed the estimation around 0 degree.

To investigate the effectiveness of the use of the likelihood distribution of pose information, we also compared the accuracy between the proposed method and a method that used a pose estimation result directly instead of using the likelihood distribution. The proposed method used likelihood values for all face directions. In contract, the comparative method used the face direction that maximized the likelihood value. Figure 7 shows the resulting estimation errors as a

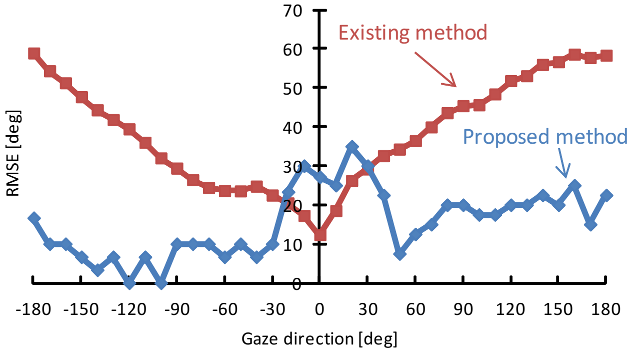


Fig. 6. Comparison of gaze estimation errors.

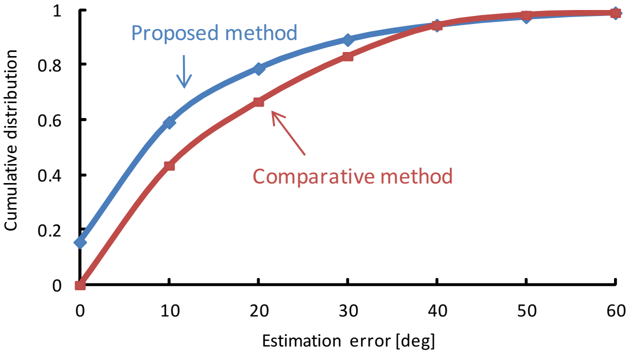


Fig. 7. Cumulative distribution of estimation errors.

cumulative distribution. From this figure, we can see that the estimation error of the proposed method is smaller than that of the comparative method. Thus we confirmed the effectiveness of the use of the likelihood distribution of pose information for gaze estimation.

V. Conclusion

We proposed a gaze estimation method using head and body pose information, and demonstrated the effectiveness of the proposed method through a gaze estimation experiment. In future work, we will use the temporal variation of the face direction as one of the pose information to further improve its accuracy of gaze estimation.

Acknowledgment

Parts of this research were supported by MEXT, Grant-in-Aid for Scientific Research. This work was developed using the MIST library (http://mist.murase.m.is.nagoya-u.ac.jp/).

References

[1] S. Ando, A. Suzuki, and H. Koike, "Measuring degree of attention to ads by appearance-based face pose estimation," in Proc. 11th Meeting on Image Recognition and Understanding (MIRU) 2008, pp.1664–1665, July 2008.

[2] A. Doshi and M. M. Trivedi, "Attention estimation by simultaneous observation of viewer and view," in Proc. 4th IEEE Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB), pp.21–27, June 2010.

[3] J. Wang and E. Sung, "Study on eye gaze estimation", IEEE Trans. Syst. Man Cybernetics Part B, vol.32, no.3, pp.332–350, June 2002.

[4] D. Beymer and M. Flickner, "Eye gaze tracking using an active stereo head," in IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR) 2003, vol.2, pp.451–458, June 2003.

[5] S. Baluja and D. Pomerleau, "Non-intrusive gaze tracking using artificial neural networks", in *Advances Neural Information Processing Systems 6, [7th NIPS Conference, Denver, Colorado, USA, 1993]*, J. D. Cowan, G. Tesauro and J. Alspector (Eds.), NIPS (Morgan Kaufmann, San Francisco, CA, USA, 1994), pp.753–760.

[6] L. Q Xu, D. Machin, and P. Sheppard, "A novel approach to real-time non-intrusive gaze finding," In Proc. British Machine Vision Conf. (BMVC) 1998, pp.428–437, Jan. 1998.

[7] K. H. Tan, D. J. Kriegman, and N. Ahuja, "Appearance-based eye gaze estimation," In Proc. 6th IEEE Workshop on Applications of Computer Vision (WACV2002), pp.191–195, Dec. 2002.

[8] L. Itti, C. Koch, and E. Niebur, "A model of saliency based visual attention for rapid scene analysis," IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI), vol.20, no.11, pp.1254–1259, Nov. 1998.

[9] A. M. Treisman and G. Gelade, "A feature-integration theory of attention," Cognitive Psychology, vol.12, no.1, pp.97–136, Jan. 1980.

[10] K. Yamada, Y. Sugano, T. Okabe, Y. Sato, A. Sugimoto, and K. Hiraki, "Attention prediction in egocentric video using motion and visual saliency," in Proc. 5th Pacific-Rim Symposium on Image and Video Technology (PSIVT) 2011, vol.1, pp. 277–288, Nov. 2011.