A Real-time Algorithm for Finger Detection in a Camera Based Finger-Friendly Interactive Board System

Ye Zhou[1] and Gerald Morrison[1]

[1]SMART Technologies Inc., Bay 2, 1440-28 Street NE
Calgary, AB T2A 7W6, Canada
{YeZhou, GeraldM}@smarttech.com

Abstract. This paper proposes an approach to finger detection for a type of camera based interactive board. In this approach, finger finding is confined within a stripe that is the projection of the edge of the board on the image plane with respect to a camera instead of using global search. The region where a finger intersects with the stripe is first detected and segmented from the background. A region growing algorithm is then applied to the region to segment the whole finger. This approach can detect multi-targets and be implemented efficiently, processing 30 or more 640×120 images per second even in a cheap DSP.

Keywords: Motion detection, Vertical intensity profile, Region of interesting, Segmentation

1 Introduction

A finger-friendly interactive board allows a user to use his/her finger as a mouse to control applications on a computer by contacting the board where the computer screen is projected. For a camera based interactive board, the core technology is a real-time surveillance system for detecting and tracking a finger. With such a system, when a finger is touching or close to the surface of the board, it is captured by digital video cameras mounted on the board. Transient images are then sent to a DSP or DSPs to detect the target, determine its contact/non-contact status with the surface and find its tip location. Many approaches to finger detection have been investigated. Hung et al [4] use an active stereo vision to detect a finger in 3D space and then track the fingertip with local search. Zhang [10] segments the finger from the background using a color-based background model. Morrison et al [5] and von Hardenberg and Bérard [2] detect the finger based on the differences between the current image and a reference image. Yang et al [9] apply a contour detection algorithm to the motion map obtained from the image differencing technique to find the hand contour and thus the fingertip.

Color-based segmentation is sensitive to the lighting condition and dependent on the quality of the digital camera. Contour detection is easily confused by background motion. Also, in those approaches based on segmenting the whole hand, it is hard to do contact detection because of the way in which the digital cameras are mounted. In

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

the approach based on the active stereo vision, it is time-consuming to detect the finger when it is first caught by the digital camera because the whole image has to be searched, and thus this approach is not suitable for the interactive board when different operations are often switched. Among the existing finger detection approaches, image differencing is the fastest and the most flexible technique. Although such an algorithm is sensitive to noise, e.g. lighting variation, the quality of the input image can be guaranteed by careful design of the mechanical structure of the interactive board and thus a high robustness can still be achieved.

In the approach of Morrison et al [5], the mechanical structure of the interactive board is designed such that four digital cameras are installed on the corners of the board to look along the board surface and a lit-bezel is mounted on the border of the screen. In this system, a finger is observed only when it is close to or touches the surface. The finger is therefore the target to be detected instead of the whole hand, which simplifies the target recognition process. The lit-bezel helps block uninteresting background motions, including shadow. The quality of input images is therefore guaranteed under most demanding lighting conditions and thus the finger detection algorithm works very robustly. The commercial products based on this approach, called DViT (www.smarttech.com), have been developed by SMART Technologies Inc.

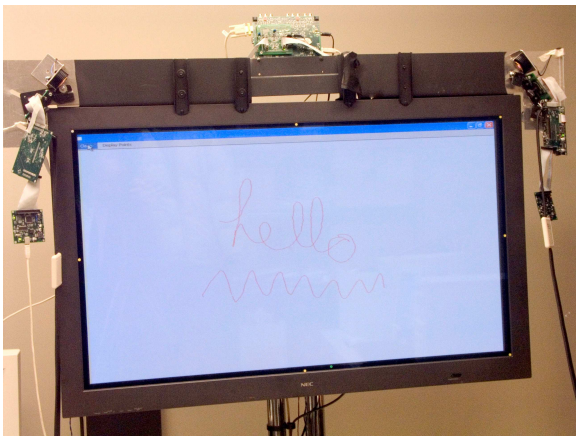

Fig. 1. Prototype of the system.

An alternative for DViT has also been created by SMART Technologies Inc. In this architecture (Fig. 1), two digital video cameras are mounted on the top left and right corners of an HDTV and two IR lights are installed right beside each camera respectively. Without the lit-bezel, the whole surface of the board and all objects close to the board and their reflections on the screen surface are within the sight of the digital cameras (Fig. 2). That means uninteresting background motion will be observed and thus finger detection in this new system becomes much more complex than in the DViT. The image processing has to eliminate not only noise caused by lighting variation, shadow, etc. but also some real moving objects, e.g. a shaking hand and its reflection on the screen surface. Moreover, the bigger input image (640×120

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8

pixels) in the new system requires faster algorithms than that used in the DViT where the input image has only 640×20 pixels.

Fig. 2. An image observed by the camera.

When a finger is detected and localized in the image captured by each camera respectively, the point on the screen surface where the finger touches can be determined by a triangulation method. This paper only discusses the image processing algorithm recognizing the finger in a single image. The rest of the paper is organized as follows. The outline of the algorithm is described in Section 2. Motion detection is discussed in Section 3. ROI (region of interesting) finding is described in Section 4. The method for finger segmentation and tip localization is presented in Section 5. Testing results on a set of typical videos are discussed in Section 6. Section 7 summarizes the algorithm.

2 Strategy of Image Processing

Due to the real-time requirement (the processing rate required is at least 30 fps and expected to achieve 120 fps), it is difficult to design an algorithm for finger detection in an image of 640×120 pixels with global search. A strategy to solve this issue is quickly finding ROIs. By analyzing the observed images (Fig. 2), two features are noticed: (1) when a finger or pen touches the surface of the screen, it must cross the edge of the HDTV in the image, which is called a "bezel"; (2) the bezel is thick enough in a 640×120 image and thus finger detection can be restricted on the bezel without losing robustness. A strategy of image processing is such determined that motion detection is restricted in a sub-image centered at the centerline of the bezel to find ROIs. When the occlusion where a finger crosses the bezel is segmented from the background, a region growing algorithm is applied to the occlusion to segment the finger from the background. The fingertip is finally localized based on the segmentation result. Since a finger has a reflection on the screen of the HDTV, contact detection can be done by checking if the finger is connected with its reflection in the image. The bezel position can be automatically detected when the digital video camera is mounted. A bezel detection algorithm does not have to be real-time as the bezel position is not changed after the camera is fixed. This algorithm is not discussed here for it is out of the scope of this paper.

3 Motion Detection

Motion detection is an essential technology for a vision system. There is a large variety of available approaches to motion detection. The most intuitive approach is to

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8

detect intensity changes. In such approaches a thresholding technique is usually employed over the difference between the current observed image and a reference image. The reference image could be the previous observed image [7] or the background image [1] [5]. In our algorithm, the background image is taken as the reference image.

The intensity change is sensitive to transient lighting variation. Sometimes such a change may even be stronger than that generated by a finger. That means that a decoy may be generated from the intensity change detection. However, the lighting variation does not affect the geometrical structure of the image in general. In other words, the boundary of an object is not changed while lighting varies. Thus, integrating the intensity change with the geometrical structure change can decrease the number of decoys and thereby increase robustness of finger detection.

The structure changes can be described using the gradient difference between the current image and the previous image. The criterion function is defined as follows:

$$\varphi(x) = \begin{cases} g_m^n(x) - g_m^{n-1}(x) & \text{if } g_m^n(x) > g_m^{n-1}(x) \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where g_m^n and g_m^{n-1} are the gradient magnitudes in the current image and the previous image respectively. It is obvious that an uninteresting structure change caused by finger's leaving will not be detected using function (1). Since a finger usually has a sharp boundary and relatively uniform intensity within the inner area and so does the bezel, using function (1) to detect the structure change on the bezel is very fast and robust.

The difference intensity image (DII) and the difference gradient image (DGI) defined in (1) are not directly used for detecting motion blobs. Instead, VIPs (Vertical Intensity Profile) of the two difference images are calculated respectively and dominant peaks in each VIP represent possible motions. That is because the difference images are very narrow and thus the VIP can enhance motions and suppress noise. Each VIP is normalized to the same range as the difference image so that the threshold for peak extraction can be used for segmentation later.

Otsu's thresholding algorithm [6] is applied to the VIPs to extract peaks. The regions that peaks locate are considered motion regions.

The background image is initialized as the first image and updated each time after the finger is found using the following criterion:

$$I_b = \alpha I_b + (1-\alpha) I_i \tag{2}$$

where I_b is the background image, $\alpha \in (0,1)$ is a scaling factor and I_i is the instant image that is defined as

$$I_i(x) = \begin{cases} \beta I_b(x) + (1-\beta) I^n(x) & x \in \text{Target} \\ I^n(x) & \text{otherwise} \end{cases} \tag{3}$$

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

where $\beta \in (0,1)$ is another scaling factor, I^n is the current image. A similar criterion was introduced by Gupte et al [1]. The difference is that in their criterion intensities at pixels in the background are not updated if the pixels are covered by a target. Our criterion still updates these pixels so that a decoy will be absorbed into background gradually.

4 ROI finding

An ROI is determined by integrating the motion regions obtained from the DII-based VIP and the DGI-based VIP. Every motion region is assigned energy that is defined as

$$E = \sum \left[dI(x) - T \right] \cdot S(dI(x) - T), \tag{4}$$

where dI is a difference image, T is the threshold for extracting the peaks, and $S(\cdot)$ is a step function:

$$S(\tau) = \begin{cases} 1 & \tau > 0 \\ 0 & \tau \le 0 \end{cases}. \tag{5}$$

These regions are then sorted in order of monotonically decreasing of energy. Only the top three dominant regions are considered as ROI related regions. If the first peak's energy is less than λE_T, where E_T denotes the total motion energy that is the sum of all the energies of the motion regions, and λ is a scaling factor (we chose λ=0.25), it is considered that energy is too scattered and thus no finger exists.

In general, the motion region generated by a finger has the highest energy. Sometimes, a sudden change of the illumination may cause a very large intensity variation that generates a motion region with higher energy than one generated by the finger (the upper picture in Fig. 3). However, the latter is usually sharper. In other words, the former can be seen as a low frequency noise. Thus, these two types of peak can be distinguished from each other in the space-Fourier domain. The finger-related peak is supposed to contain most dominant frequencies from low to high. The illumination-change-related peak contains only lower dominant frequencies. This fact can be observed by means of the multi-resolution time-frequency analysis (the middle picture in Fig. 3) [8]. Since the time-frequency analysis is a time-consuming method, low frequency noise is in practice suppressed by a band pass filter. The two cutoff frequencies can be predetermined based on experience.

The ROI is chosen from the DII-based motion regions, and the DGI-based motion regions help eliminate decoys. A DII-based motion region is considered as a candidate of the ROI if and only if there are DGI-based motions around its boundaries or it overlaps the finger position in the previous frame. The candidate with the highest energy is then considered as the current ROI.

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

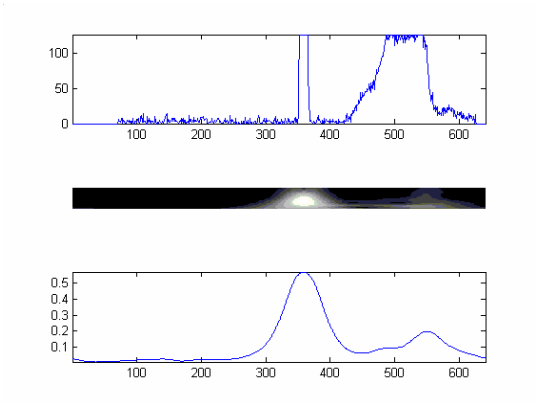

Fig. 3. Top: a VIP of a difference image; middle: local frequency spectra of the VIP; bottom: the VIP of the spectral map.

An example of ROI finding is shown in Fig. 4. There are two moving objects in the top image: one is a finger touching the surface and its reflection; the other is a hand moving beside the board and its reflection that are uninteresting movements and thus eliminated. The middle image shows the VIP of the DII. It is noticed that the hostile movement also generates strong peaks. The bottom image gives the location of the ROI.

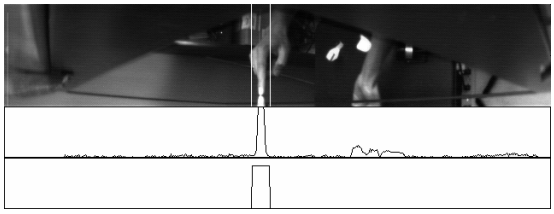

Fig. 4. Top: a finger and its reflection; middle: the VIP of the DII; bottom: the ROI.

5 Finger Segmentation and Tip Localization

Once the ROI is determined, the finger can be segmented from the background. At first, a seed region is obtained by thresholding the difference image bounded by the ROI. The seed region is then grown within the ROI using a criterion inspired from Hojjatoleslami and Kittler's algorithm [3]. The result of the segmentation is the occlusion where the finger intersects with the bezel. Finally, the occlusion is grown upwards (because the finger comes from the lower part in the image) row by row using the same criterion until the stopping condition is satisfied.

The criterion for region growing is based on two concepts proposed by Hojjatoleslami and Kittler [3]. One is called the average contrast boundary of the

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

region, which is defined as the difference between the average intensity of the region and the average intensity of the current boundary, i.e.

$$b_{AC} = \left| \frac{1}{N_R} \sum_{x \in \mathbf{R}} I(x) - \frac{1}{N_C} \sum_{x \in \mathbf{B}_C} I(x) \right| \tag{6}$$

where $\mathbf{R}$ is the current region, N_R is the number of pixels in the current region, $\mathbf{B}_C$ is the current boundary of the region, and N_C is the number of pixels at the current boundary. The other is called the peripheral contrast boundary of the region, which is originally defined as the difference between the average intensity of the current internal boundary and the average intensity of the current boundary. The so-called internal boundary is the set of outermost pixels within the current region. The concept is modified in this paper and redefined as the difference between the average intensity of the current internal boundary and the average intensity of the current external boundary, i.e.,

$$b_{PC} = \left| \frac{1}{N_I} \sum_{x \in \mathbf{B}_I} I(x) - \frac{1}{N_E} \sum_{x \in \mathbf{B}_E} I(x) \right| \tag{7}$$

where $\mathbf{B}_I$ is the current internal boundary, N_I is the number of pixels at the current internal boundary, $\mathbf{B}_E$ is the current external boundary, and N_E is the number of pixels at the current external boundary, which is defined as the set of innermost pixels out of the current region including the current boundary. The region grows if both b_{AC} and b_{PC} do not decrease. Within the ROI, it only grows leftwards and rightwards. The growing process stops when one of the following three conditions is satisfied:

(1) the region cannot grow according to the growing criterion;
(2) the region has grown for M_r rows where M_r is the maximum height of the finger, which is the maximum distance from the bezel to the fingertip and determined according to the width of the occlusion;
(3) the region reaches the upper boundary of the display area.

When the finger is segmented from the background, its left and right boundaries are also obtained. Let $b_L(r)$ and $b_R(r)$ represent the left boundary and the right boundary at the r-th row respectively. Assume that the region starts at row r_s and ends at row r_e. The set of fingertip candidates can be denoted as

$$\Omega = \left\{ (r,c) \middle| c = \frac{b_L(r) + b_R(r)}{2}, r = r_s, r_s + 1, \ldots r_e \right\}.$$

An energy function is defined to find the fingertip position:

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

$$E(r,c)=\min\left\{\left|\frac{1}{c-b_L(r)+1}\sum_{t=b_L(r)}^{c}\sum_{s=r-\Delta r}^{r}\left[I(s,t)-I(r,t)\right]\right|,\ \left|\frac{1}{b_R(r)-c+1}\sum_{t=c}^{b_R(r)}\sum_{s=r-\Delta r}^{r}\left[I(s,t)-I(r,t)\right]\right|\right\}, \tag{8}$$

where Δr defines a row-neighborhood. This definition is based on the fact that pixels within a finger are brighter than those at boundary area of the finger under IR lighting. Thus, intensity difference between pixels at the row where tip is located and pixels at neighboring rows should be greater than intensity difference between pixels at another row covered by the finger and pixels at its neighboring rows. In other words energy function (8) reaches the maximum at the fingertip position (r_t, c_t), i.e.

$$(r_t, c_t)=\arg\max_{(r,c)\in\Omega}\{E(r,c)\}. \tag{9}$$

Three typical cases of the finger detection and tip localization are given in Fig. 5. Fig. 5(top) shows that a finger is coming but not closes to the surface because it has not crossed the bezel. It is thereby ignored. In Fig. 5(middle) the finger has crossed the bezel. That means it is very close to the surface. Since the finger does not link with its reflection, it has not contacted the surface and therefore region growing stops at the position of the fingertip. In Fig. 5(bottom) the finger has touched the surface. The region growing stops at the maximum height of the finger and thus the result of the segmentation covers the finger and part of its reflection.

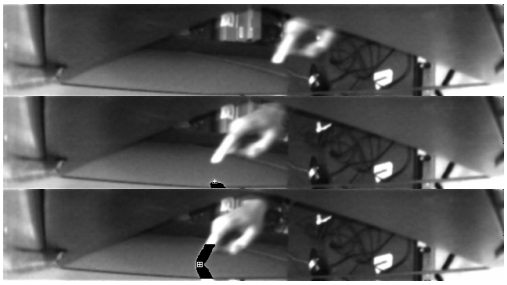

Fig. 5. Top: a finger is coming but has not crossed the bezel yet; middle: the finger crosses the bezel but does not touch the surface; bottom: the finger touches the surface.

6 Experimental Results

The algorithm has been implemented in DSPs and works very well on a series of on-line testing on the prototype. Table 1 gives a set of off-line testing results on eight typical videos. The target is a pen in video 1 and a finger in all other videos. Videos 2

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

and 3 were collected for testing low frequency noise elimination. Hot spots were generated in these videos by an IR light. In video 2, the hot spots were still and finger detection was not disturbed. When collecting videos 3, we moved the IR light and generated some moving hot spots. Because the boundary of the bezel disappeared under strong IR lighting, some hot spots on the bezel were linked together with background objects. Ten moving hot spots were misrecognized as targets. This type of decoy may be eliminated in the next stage of triangulation if no corresponding decoy is generated in another camera or the triangulating position is outside the displaying area. Video 4 was used to test the algorithm's performance in a dark environment. Video 5 was collected for testing a finger clicking the board. Video 6 is for testing big finger recognition. Because the hand was very close to the camera, knuckles also crossed the bezel and thus were detected as targets. In addition, knuckles may also affect finger recognition when their motion energy is stronger than finger's. Fortunately, when a finger is close to one camera, it must be far away from another camera. The knuckles should not be observed by the other camera unless the finger is too skewed. Therefore, such a decoy can be eliminated in the stage of triangulation. Video 7 provided a case study for detecting a blurred and skewed finger. Background motion is also a common issue in finger detection. Such a motion is not interesting and may interfere detecting an interesting motion. Video 8 was created for testing the capability of the algorithm to detect finger when disturbed by a background motion.

Table 1. Off-line text results on 8 typical videos.

Video	Frames	Targets	Detected	Decoys	Recognition rate (%)
1	201	163	163	0	100
2	306	166	166	0	100
3	200	160	160	10	100
4	205	131	131	0	100
5	178	150	149	0	99.33
6	247	147	143	73	97.28
7	213	158	157	0	99.37
8	77	76	76	0	100
Total	1627	1151	1145	83	99.48

7 Conclusions

In this work, a framework has been proposed for detecting a finger and finding its tip position for a type of passive vision system. According to this framework, a very simple interactive board can be built by integrating an HDTV, two cameras with IR assist lights, and DSPs. The two cameras have to be mounted at a proper height based on two criteria:

(1) The bezel in the observed image must be at least three pixels wide.

(2) A finger crosses the bezel in the image only when it is very close to the surface so that knuckle or palm does not intersect the bezel when the finger is not too skewed.

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

For reflections on the bezel are much weaker than on the board surface, the first criterion is very effective to reduce the influence of background objects' reflections on the motion detection in the sub-image. The second criterion simplifies the finger segmentation and tip localization.

A finger can be quickly detected and localized by confining motion detection within the sub-image round the bezel. In addition noise and uninteresting background motions can be eliminated to the utmost. Using VIP instead of pixelwise differences to define motion regions can increase robustness of motion detection. The band pass filter has to be carefully designed so that detecting a big finger is not affected while low frequency noise is suppressed.

Although the algorithm is designed for finger detection, it also works for pen detection if the pen tip and the penholder have the same color. Besides HDTV based interactive boards, the algorithm can also be used in a rear projection system and may be extended to a front projection system.

Acknowledgement. Many thanks to Dr. David Holmgren in SMART Technologies Inc. for his help in our organizing this paper.

References

1. Gupte, S., Masoud, O., Martin, R. F. K., and Papanikolopoulos, N. P.: Detection and classification of vehicles. IEEE Transactions on Intelligent Transportation Systems, Vol. 3, No. 1, (2002) 37-47
2. von Hardenberg, C. and Bérard, F.: Bare-hand Human-computer Interface. In Proceedings of the ACM Workshop on Perceptive User Interfaces, Orlando, Florida, USA, Nov (2001) 15-16
3. Hojjatoleslami, S. A., and Kittler, J.: Region growing: a new approach. IEEE Transactions on Image Processing, Vol. 7, No. 7, (1998) 1079-1084
4. Hung, Y. P., Yang, Y. S., Chen, Y. S., Hsieh, I. B., and Fuh C. S.: Free-Hand Pointer by Use of an Active Stereo Vision System. In Proceedings of the 14th International Conference on Pattern Recognition, Brisbane, Australia, Aug. Vol. 2, (1998) 1244-1246
5. Morrison, G., Singh, M., and Holmgren, D.: Machine vision passive touch technology for interactive displays. SID 2001 Tech. Dig. 3, (2001) 74-77
6. Otsu, N.: A threshold selection method from gray level histograms. IEEE Transactions on Systems, Man and Cybernetics. SMC-9, (1979) 62-66
7. Paragios, N. and Deriche, R.: Geodesic active contours and level sets for the detection and tracking of moving objects. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 3, (2000) 266-280
8. Stockwell R G., Mansiha, L. and Lowe R. P.: Localization of the complex spectrum. IEEE transactions on Signal Processing. Vol. 44, No. 4, (1996) 998—1001
9. Yang, D. D., Jin, L. W., Yin, J. X., Zhen, L. X., and Huang J. C.: An Effective Robust Fingertip Detection Method for Finger Writing Character Recognition System, In Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, China, August (2005) 18-21
10. Zhang, Z.: Vision-based interaction with fingers and papers. In Proceedings of International Symposium on the CREST Digital Archiving Project, Tokyo, Japan, May (2003) 23-24