Virtual Reality manuscript No.
(will be inserted by the editor)

Document Search Support by Making Physical Documents Transparent in Projection-Based Mixed Reality

Daisuke Iwai · Kosuke Sato

Received: date / Accepted: date

Abstract This paper presents *Limpid Desk*, which supports document search on a physical desktop by making the upper layer of a document stack transparent in a projection-based mixed reality environment. A user can visually access a lower-layer document without physically removing the upper documents. This is accomplished by superimposition of cover textures of lower-layer documents on the upper documents by projected imagery. This paper introduces a method of generating projection images that make physical documents transparent. Furthermore, a touch sensing method based on thermal image processing is proposed for the system's input interface. Areas touched by a user on physical documents can be detected without any user-worn or handheld devices. This interface allows a user to select a stack to be made transparent by a simple touch gesture. Three document search support techniques are realized using the system. User studies are conducted and the results show the effectiveness of the proposed techniques.

Keywords Projection-based mixed reality · Document search support · Making documents transparent · Thermal image processing · Thermal trace · Touch sensing

1 Introduction and Motivation

Physical documents (e.g., magazines, books, and papers) are usually scattered and piled chaotically on a desktop. The disorder on the desktop makes searching for a desired document burdensome. When all the documents are organized, the desired document can easily be found, and consequently, work can be handled efficiently. However, most people are generally too busy to clean up their desktops.

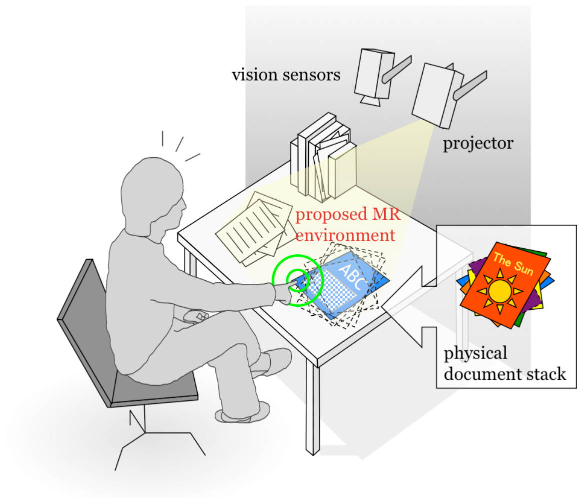


Fig. 1 Concept of *Limpid Desk*

Several window management tools have been developed with which either windows or file icons can be efficiently searched on a personal computer (PC) desktop. For example, *Exposé* [1] allows a user to quickly locate an open window without the need to click through many windows to find a specific target. In particular, the *All windows* feature shows all open windows, shrinking their appearance so that they all fit on a single screen. *Windows 7* [2] has a feature where all the open

D. Iwai · K. Sato
Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan
Tel.: +81-6-6850-6372
Fax: +81-6-6850-6341
E-mail: {daisuke.iwai, sato}@sys.es.osaka-u.ac.jp

[1] Apple Inc., `http://www.apple.com/`
[2] Microsoft Corporation, `http://www.microsoft.com/`

windows are made transparent to make file icons on a desktop visible. These tools can be activated by a single action, such as pressing a key or moving the mouse cursor to a corner of the desktop.

This paper presents a novel document search support system in which documents are made transparent in situ on a physical desktop. Usually the upper-layer documents in a stack are opaque and users have to remove them to look at lower-layer documents. If the upper-layer documents become transparent, the lower-layer document is visible. In such an environment, users can easily find a specific target in a document stack without removing the upper documents.

Limpid Desk uses projection-based mixed reality (MR) technology to change the appearances of physical documents so that they look transparent in real space (Fig. 1). First, the system combines surface textures of the lower-layer documents and the desktop under the upper-layer documents to be made transparent. Then geometric registration and radiometric compensation are applied to the image. The generated image is projected onto the upper-layer documents with additional visual effects so that a user can feel that the documents are made transparent and easily count the number of documents that are already made transparent. These effects allow the user to remove a target lower-layer document from the stack immediately after the target document is virtually exposed. A novel touch sensing method using thermal image processing is also proposed. Areas that a user touches on physical documents can be detected without any user-worn or handheld devices.

Based on these basic technologies, this paper proposes three document search support techniques. A user can search for a desired document by directly touching a document stack in real space. Conventional interfaces, such as computer monitors, keyboards, and mice, are not needed. Because all the interaction spaces are unified on a physical desktop, an efficient document search is realized.

The remainder of this paper is organized as follows. Related studies are briefly described in the subsequent section. Detailed information about how to make physical documents transparent and the touch sensing method is provided in Sec. 3 and 4, respectively. System configuration, document search support techniques, and user studies are presented in Sec. 5. The advantages and limitations of the proposed approach are discussed in Sec. 6. Finally, the conclusion of the paper is provided in Sec. 7.

2 Related Work

Projection-based MR focuses on changing attributes (e.g., color, texture, and shape) of a physical object by projected imagery (Bimber and Raskar, 2005). In recent years, this has been exploited for making physical objects transparent (Inami et al, 2003; Bonanni et al, 2005). However, these objects were limited to those suitable for image projection, such as a retro-reflective material and a refrigerator's simple white door. Physical documents, such as magazines and photos, have spatially varying reflectance properties on their surfaces, which disturb the appearance of projected images.

Several researchers focus on controlling a physical object's appearance by a projection image even when the object has complex reflectance properties. This can be achieved by a radiometric compensation process (Bimber et al, 2008). We apply one of these techniques (Yoshida et al, 2003) to make physical documents with complex textures transparent.

Previous studies try to support a user to search a physical document with a PC interface. Kim et al. proposed tracking physical documents over time using an overhead camera (Kim et al, 2004). The tracked documents are automatically linked to the corresponding electronic documents, which are stored on a PC. The user can easily locate the desired document in physical stacks by performing a keyword search. *Strata Drawer* supports document search in a physical drawer (Siio et al, 2003). A camera is mounted facing downward in the drawer. When a user places a document in it, a photograph is automatically taken. Then the user can browse pictures of documents in the drawer with PC interfaces.

In these systems, a user has to move to a place where the desired document exists after searching for it on a PC. A spatial seam exists between the manipulation-and-display space (PC) and the search space (desktop). In contrast, *Limpid Desk* is designed to unify these spaces on a physical desktop to realize more efficient document search. In addition, we assume that documents targeted by *Limpid Desk* include those that cannot be located by a keyword search, such as photos or handwritten notes.

A touching action is used as a trigger operation in *Limpid Desk* to support a user's direct manipulation of physical documents. Several methods have been developed to measure user actions, such as a magnetic tracker (Bandyopadhyay et al, 2001), sensors embedded in the projection surface (Dietz and Leigh, 2001), and vision-based finger gesture recognition (Koike et al, 2001). In the first solution, a user has to hold a magnetic tracker to interact with the system. The device interrupts the natural document search interaction. The second solution is unsuitable for the proposed system because it cannot detect a user's touch actions on a physical document stack. The vision-based solution can robustly recognize a user's finger gestures. However, it also has difficulties with detecting a user's touch action. In contrast, Wilson proposed using the shadows of a user's fingers to detect touch points on the surface (Wilson, 2005). But this shadow-based algorithm cannot work well on a desktop where various objects are scattered. This paper proposes a novel touch sensing method using thermal images. In this method, touched areas can be detected on physical documents that are scattered and piled chaotically on a desktop. This approach is developed on the basis of a previous touch-detection method where areas that a user touches can be detected only on a thin paper screen (Iwai and Sato, 2005).

The basic concept of *Limpid Desk* and several related technological principles were preliminarily proposed in short reports (Iwai et al, 2006; Iwai and Sato, 2006). This paper extends these technologies and furthermore conducts user studies to confirm the effectiveness of the proposed document search support techniques.

3 Making Physical Documents Transparent in Projection-Based MR

This section introduces a fundamental optical model of making a physical document transparent by projected imagery. Additional visual effects are also designed specifically to support document search.

3.1 Fundamental Optical Model

We apply a simple optical model to describe the appearance of a document stack where the upper-layer documents are transparent. A radiometric compensation technique is also applied to display the desired appearance on the textured surfaces of the documents. A proof-of-concept experiment is conducted to confirm the validity of the model.

3.1.1 Principle

This paper applies the alpha-blending approach to realize virtually transparent documents (Fig. 2). Although the proposed model does not follow real physics of reflection and transmission, we found that it works quite well to display transparent effect. First, the appearance of a transparent document (spectral reflectance $M_1(\lambda)$,

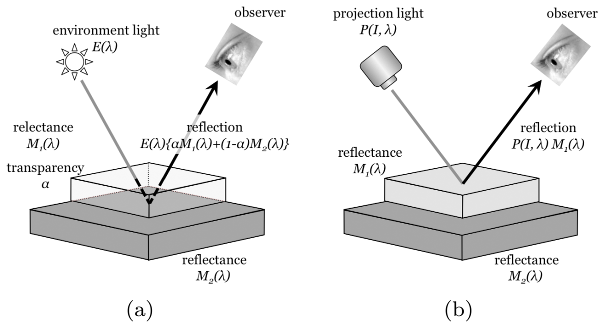


Fig. 2 Optical model: (a) upper document is transparent under environment light and (b) upper document is opaque under projection light

transparency α) that covers an opaque document (spectral reflectance $M_2(\lambda)$) under environment light $E(\lambda)$ is considered (Fig. 2(a)). The observed spectral distribution of the reflection $R(\lambda)$ is computed by

$$R(\lambda) = E(\lambda)\{\alpha M_1(\lambda) + (1 - \alpha)M_2(\lambda)\}, \quad (1)$$

where λ represents the wavelength of light. $R(\lambda)$ is the desired appearance to be displayed on the surface by projection-based superimposition.

Second, the *Limpid Desk* configuration is considered. In this case, the environment light is replaced with a projector and the upper document is opaque. The observed reflection $R(\lambda)$ of a projection light $P(\lambda|I_R, I_G, I_B)$ is computed by

$$R(\lambda) = P(\lambda|I_R, I_G, I_B)M_1(\lambda), \quad (2)$$

where I_R, I_G, I_B are the input intensity values of the projector ($0 \leq I_R, I_G, I_B \leq 1$). $P(\lambda|I_R, I_G, I_B)$ contains both the projector-to-surface form-factor and the black-level of the projector. The fraction of light that arrives at a surface patch depends on the geometric relationship between the light source and the surface, particularly the angle between the light ray and the surface normal, and the distance between the light source and the surface. The form-factor is the simplest way to approximate this fraction. The black-level represents the projector's minimum intensity level.

As the environment light of the system, the projector illuminates the desktop with half of the maximum intensity:

$$E(\lambda) = P(\lambda|0.5, 0.5, 0.5). \quad (3)$$

This was chosen to avoid clipping artifacts in the following radiometric compensation process. Because a dim environment light makes it difficult for users to read documents, the light should be as bright as possible. However, if the maximum intensity of the projector is applied as the environment light, correct colors cannot be displayed when the system tries to display a brighter color than the reflectance color.

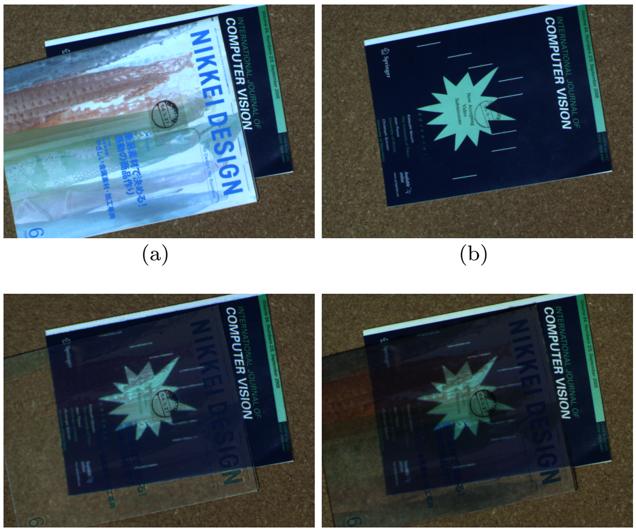
Every time a new document is placed on the stack, its appearance under the environment light $E(\lambda)M(\lambda)$ is captured by an overhead visible camera. The target appearance can be calculated from the stored appearances on the basis of Equation 1. Once the target appearance is defined, the per-pixel projection color is calculated through a radiometric compensation technique that we previously proposed (Yoshida et al, 2003). The detailed principle of the radiometric compensation technique is described in Appendix A. Because the method cancels complex textures on the projection surface, a better transparent appearance is realized.

The applied radiometric compensation requires color and geometric calibration processes. The overhead visible camera which captures the reflection of the projection light is used for them. In color calibration, uniform colored patterns are projected to calculate color-mixing matrices for each camera pixel, which are required for radiometric compensation.

In geometric calibration, structured light patterns (particularly graycode patterns) are projected onto the document stack to obtain camera-to-projector pixel correspondences (*C2P map*) (Sato and Inokuchi, 1987). If intrinsic and extrinsic parameters of the visible camera and the projector are calibrated, 3-D shape of the document stack can be measured through the C2P map. Consequently, the desired appearance can be rendered for an arbitrarily viewpoint by a 2-pass rendering method described in (Bimber and Raskar, 2005). In the following experiments and pilot study, we fix the user's viewpoint. In a practical application, an eye tracking sensor can be integrated into the proposed system.

3.1.2 Proof-of-Concept Experiment

A fundamental experiment was conducted to confirm the effectiveness of the proposed principle. Two documents were placed one by one on a desktop under the environment light provided by the projector, as shown in Fig. 3(a). When each document was placed, a color image was captured to obtain $E(\lambda)M_1(\lambda)$ and $E(\lambda)M_2(\lambda)$. We calculated the target appearance by using Equation (1) with α (i.e., the transparency of the upper-layer document) of 1.0 (Fig. 3(b)). Note that $R(\lambda)$ represents the target appearance in this case. Therefore, the color of the upper-layer document does not remain in the target appearance. Figure 3(c) shows a projection result based on the proposed principle. It is confirmed that the target appearance is almost reproduced on the textured surface. Figure 3(d) shows the result without radiometric compensation. We confirmed that the radiometric compensation technique improved the appearance of transparency.



(a) (b) (c) (d)

Fig. 3 Results of fundamental experiment: (a) a document stack under projected environment light, (b) target appearance, (c) projection result, and (d) projection result without radiometric compensation

3.1.3 Limitation

A major drawback of the current implementation is that geometric and radiometric calibration needs to be done every time there is a change in the scene. For the calibration, multiple pattern lights have to be projected. Because it takes at least a few seconds, this process disturbs a user's natural document search process. On the other hand, several researchers have focused on fast and imperceptible calibration methods (Zollmann and Bimber, 2007). We believe that such a method can solve this problem.

Limited dynamic range of projection is another issue. In some cases, real documents cannot be completely made transparent due to this limitation. Several researchers have attempted to solve the issue by adjusting the desired appearance before radiometric compensation is carried out (Grundhöfer and Bimber, 2008). This reduces perceived visual artifacts while simultaneously preserving a maximum of luminance and contrast. Furthermore, high dynamic range projection technologies have also been developed in research fields of computer graphics and projection display (Seetzen et al, 2004). We believe that these technologies can solve the issue.

3.2 Additional Visual Effects

In addition to the fundamental optical model, two types of visual effects are applied. As shown in Fig. 3, the system can change the appearance of the upper-layer document so that the lower-layer document becomes visible to the user. However, the user may not feel that the document is made transparent because the visualization is unrealistic. Thus, one of the effects is designed to allow a user to feel that the upper-layer document becomes transparent.

Because the purpose of the system is to support document search, users must be able to pick up a desired document immediately after it is visually exposed. Therefore, the other effect is designed to allow users know how many documents are already made transparent.

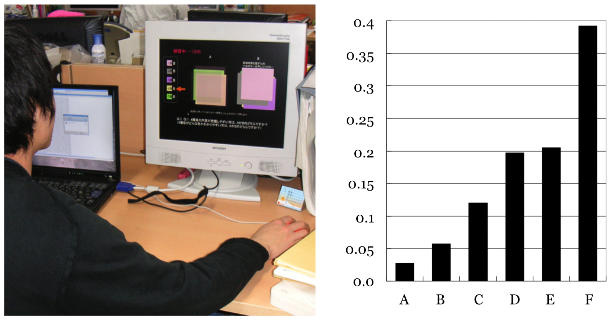
3.2.1 Enhancement of Transparent Effect

Fadeout is applied to the first visual effect, which allows a user to feel that the upper-layer document becomes transparent. This visual effect gradually reduces the opacity of the upper-layer document.

A questionnaire survey was conducted to evaluate the effect. Seven participants were recruited from a local university. In the experiment, two papers were piled on a desktop. The participants saw that the upper one was made transparent with the fadeout effect by projected imagery. They also saw that the upper one was made transparent without the effect. All participants answered that they were able to feel that the upper-layer document was made transparent with the effect more than without the effect. The result indicates the effectiveness of the proposed effect.

3.2.2 Indication of Layer Number of Visually Exposed Document

In document search scenarios, it is important for users to comprehend the layer number of the visually exposed lower-layer document. This allows users to take a document from the stack immediately after it is exposed. It is important to provide users with information about the number of upper-layer documents that are already made transparent. A straightforward solution is alpha blending of all the invisible documents. However, simple alpha blending considerably decreases readability of the blended image contents because they visually interfere with each other (Baudisch and Gutwin, 2004). Instead, we propose an effect that makes upper-layer documents transparent one by one while retaining all disappeared documents' frames.



(a) (b)

Fig. 4 AHP to validate the proposed visual effect: (a) view of the experiment and (b) result

The analytic hierarchy process (AHP) is conducted to prove that the proposed effect is efficient. Six visual effects (A, B, ..., F) including the proposed one are compared. Effect A makes the upper-layer documents transparent without any additional effects. Effect B makes them transparent with alpha blending. Effect C makes them transparent while retaining all disappeared documents' frames. In these effects, the upper-layer documents disappear at once. Effects D, E, and F are the same as A, B and C, respectively, except that they make the upper-layer documents transparent one by one. Thus, effect F is the one that is proposed.

Fifteen participants were recruited from a local university. Each pair of objective visual effects was displayed on a PC monitor (Fig. 4(a)). The participants saw the effects and compared them. They assigned 1 to the effect that makes it easier to understand the layer number of the exposed lower-layer document and 0 for the other effect of each pair. The result shows the relative scores of the compared effects (Fig. 4(b)). This result indicates that the proposed effect is effective for helping a user understand the layer number of the visually exposed lower-layer document.

3.2.3 Combination of Two Visual Effects

The aforementioned visual effects are combined and implemented in the proposed system. The combined effect is shown in Fig. 5 where the upper-layer documents are gradually made transparent one by one while retaining the frames of the disappeared documents.

4 Touched Area Sensing

This section describes a touch sensing method using thermography. The proposed method can detect an instantaneous touch action on a physical document placed at various heights on the desktop. In general, when a hand touches an object, body heat is transferred to the

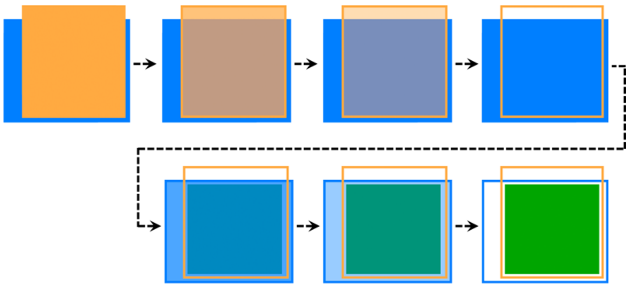


Fig. 5 Example of combination of two visual effects

object. After the hand releases the object, the body heat stays for a while. This residual body heat is called a *thermal trace* in this paper. The proposed method extracts the thermal trace of a user's instantaneous touch on a physical document. Note that the following image processing is performed on 8-bit images.

4.1 Method

An overhead thermo-infrared (IR) camera is used to capture the temperature distribution of a desktop scene. A warm area in the captured IR image is detected as a touched area through a thresholding process. However, besides the touched area, other warm objects such as a user's hand and a coffee cup on the desktop are also detected. Therefore, it is necessary to detect only the user's touch area from warm areas. To achieve this, an additional visible camera is used, which captures the scene from the same perspective as the IR camera.

We assume that the thermal trace caused by a user's instantaneous touch action is static in the visible image and dynamic in the IR image. If an area in an image does not change for at least ten seconds, it is regarded as static; it is considered dynamic otherwise. Warm areas are classified into four categories, as shown in Table 1.

Case 1: A warm area that is static in both the visible and IR images (e.g., a coffee cup or a user's hand that is placed on the desktop or a document and does not move for more than ten seconds)

Case 2: A warm area that is static in the visible image and dynamic in the IR image (a touched area based on our assumption)

Case 3: A warm area that is dynamic in the visible image and static in the IR image (e.g., a small TV screen that is placed on the desktop)

Case 4: A warm area that is dynamic in both the visible and IR images (e.g., a coffee cup moved by a user, a moving hand, or a thermal trace caused when an object of case 1, such as a static coffee cup, is released from the desktop or the document)

Table 1 Categories of warm areas

case	visible image	IR image		touch area
1	static	static	$\Rightarrow$	no
2	static	dynamic	$\Rightarrow$	yes
3	dynamic	static	$\Rightarrow$	no
4	dynamic	dynamic	$\Rightarrow$	no

Figure 6 shows the process flow of the proposed method. This extracts only a warm area of case 2, which is static in the visible image and dynamic in the IR image. First, dynamic areas in the visible image such as a user's moving hand are removed by foreground subtraction followed by a mask process. Second, warm areas are extracted from the masked IR image by a thresholding process. Third, background subtraction is applied to the IR image to remove static warm areas such as a coffee cup on the desktop. Finally, an additional noise cancellation process enables detection of only the touched area. This process is performed at 15 frames per second with an experimental system that will be introduced in the following section. Note that the proposed method updates background images by taking new visible and thermal images when there is no change in the captured scene for a certain time (ten seconds).

This technique can selectively detect a warm area of case 2 as shown in Table 1. However, if a warm object such as a coffee mug is placed on a desktop for a couple of seconds and then moved away from the desktop, it belongs to case 2 and is detected as a user's hand touch. The failure can be solved by applying an image-based object recognition technique to distinguish a user's hand from other objects (Swain and Ballard, 1991).

4.2 Investigation of Thermal Trace Transition

Heat conduction is generally considered a slow phenomenon. We investigate the transition of thermal traces on a document. The transitions on copper and thermoplastic acrylonitrile butadiene styrene (ABS) have been studied (Ho et al, 2007). However, to our knowledge, such previous research has not well investigated the thermal trace transition on paper.

We evaluate how long it takes to transfer body heat from a user to a document. They also evaluate how long the heat stays on the touched area after the hand is released. A piece of paper is used as a touch surface. The paper is thin enough so that the same temperature distribution occurs on both sides of the paper simultaneously. In the experiment, a hand touches the paper for about one second. The thermo-IR camera measures

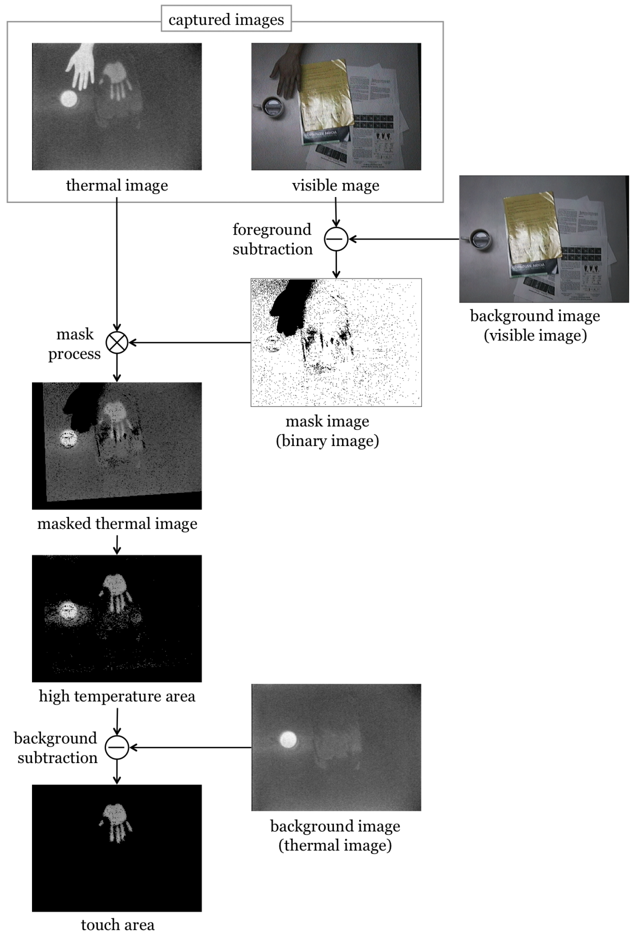


Fig. 6 Process flow of touch sensing method

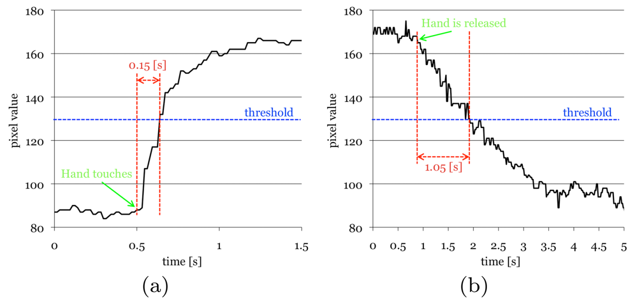


Fig. 7 Thermal transition on paper with one second of hand contact: (a) hand touches the paper and (b) it is released

the temperature transition of the touched area from the other side of the paper.

Figure 7(a) shows that the temperature reaches the predetermined threshold (pixel value: 130) from room temperature (pixel value: 90) in 0.15 [s] after the hand touches the paper. This result confirms that the transfer time is short enough so that a user can manipulate the system by a daily touching action. Figure 7(b) shows that the temperature falls below the threshold 1.05 [s] after the hand is released from the paper. This result confirms that the user has to wait for about one second to perform another manipulation on the same document stack. Considering this transition time, we designed the document search support techniques introduced in the following section.

5 Document Search Support System

This section describes how the proposed document search support system works. First, the proposed system configuration is introduced. Then three document search support techniques are proposed. Finally, user studies are conducted to evaluate how the proposed techniques improve document search.

5.1 System Configuration

Figure 8 shows a view of the experimental setup of *Limpid Desk*. The system consists of a video projector, two visible cameras, and a thermo-IR camera, which are mounted overhead and facing the desktop. This system runs on a PC (Intel Pentium-4 2.5 [GHz]). A video projector (NEC MT1075J) renders document images directly on the stacks. As explained in 3.1, this is also used as the environment light on the scene. A visible camera (Point Grey Research Scorpion) which is next to the projector is used for geometric and radiometric image correction. Note that gamma characteristics of the devices are already corrected.

Another visible camera (Sony EVI-G20) is used for the proposed touched area sensing. A thermo-IR camera (Mitsubishi IR-SC1) takes a thermal image that is also used for touched area sensing. The optics of the thermo-IR camera and the visible camera are coaxial. A dichroic mirror is used, which passes visible light and reflects thermo-IR light. Images are geometrically registered via homography, as proposed in a previous study (Yasuda et al, 2004). This setup allows users to move their hands freely on and over the desktop.

5.2 Document Search Support Techniques

We implement three different document search support techniques in which a user can see through stacked documents only by touching them without any user-worn or handheld devices.

In the first technique, after a user touches the top of a document stack, documents in the stack are made transparent one by one to the bottom (Fig. 9). The documents disappear according to the visual effect proposed in 3.2. As a result, the user can browse all the documents in the stack by one direct manipulation without

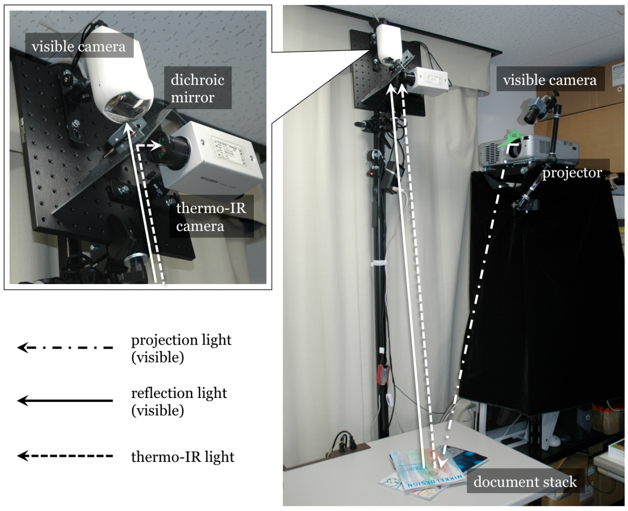


Fig. 8 View of experimental setup

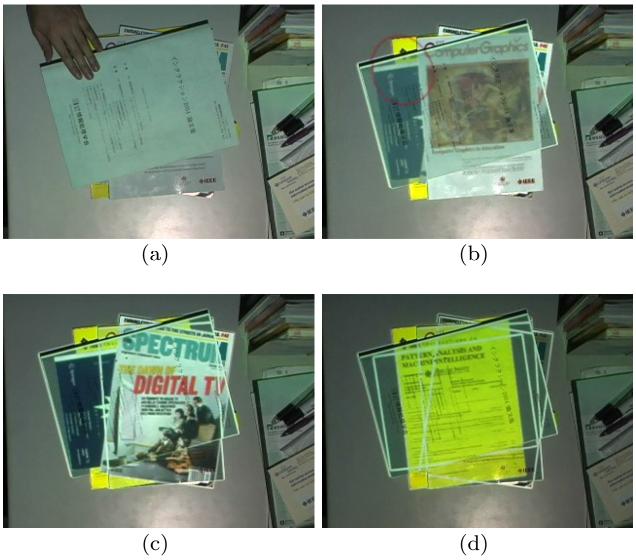


Fig. 9 Browse Search: (a) a user touches the top document, (b) the touched area is detected, (c) documents are gradually made transparent, and (d) the bottom document is visually exposed

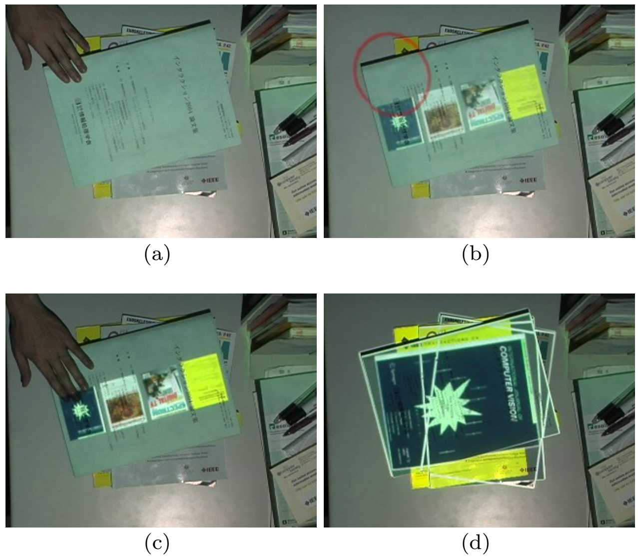


Fig. 10 Thumbnail Search: (a) a user touches the top document, (b) thumbnail images are exposed, (c) user touches one of the thumbnails, and (d) the selected document is exposed

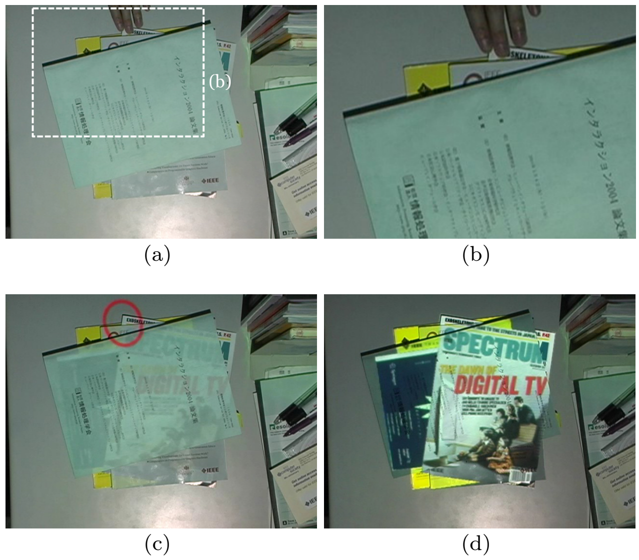


Fig. 11 Direct Select: (a) a user touches a document, (b) magnified view, (c) the touched area is detected, and (d) the selected document is exposed

physically removing the upper layer documents. Furthermore, if the desired document is found, the user can immediately take it from the stack as a result of the visual effect. The technique is referred to as **Browse Search**.

In the second technique, after a user touches the top of a document stack, thumbnail images of all documents in the stack are aligned and projected onto the top (Fig. 10). If the desired document appears, the user touches its thumbnail image. Then the overlying documents are made transparent according to the visual effect proposed in 3.2. The technique allows the user to comprehend all the documents in the stack at a glance. The technique is referred to as **Thumbnail Search**.

In the third technique, a user can directly select a desired document by touching its edge (Fig. 11). All the upper layer documents disappear and the target document is exposed immediately after the user touches its edge. The technique is referred to as **Direct Selection**.

The first two techniques are designed to be used in a scenario where a user searches for a document by seeing all documents in a stack. On the other hand, the last technique is designed to be used not in a document search scenario, but in a scenario where the user is interested in a document in a stack and would like to quickly check what it is.

Table 2 Measured document search time

measured time	document search technique	target document
$t_n(x\|HS)$	Hand Search	not exist
$t_e(x\|HS)$	Hand Search	exist
$t_n(x\|BS)$	Browse Search	not exist
$t_e(x\|BS)$	Browse Search	exist
$t_n(x\|TS)$	Thumbnail Search	not exist
$t_e(x\|TS)$	Thumbnail Search	exist

5.3 Evaluation of Browse Search and Thumbnail Search

A user study was conducted to address the issue of whether or not the proposed document search techniques are any better than searching for documents by hand in terms of document search time. Document search only by hand is referred to as **Hand Search**. We compared the first two interaction techniques, Browse Search and Thumbnail Search, with Hand Search.

5.3.1 Document Search Time Model

First, we propose a model of document search time. In general, multiple document stacks are placed on a desktop. Therefore, it is more reasonable to consider a situation in which a person searches for a desired document from multiple stacks rather than from a single stack. When the person searches from L document stacks and finds it in the l-th stack (i.e., $l < L$), the total amount of document search time $T(l|L)$ is modeled as:

$$T(l|L) = \begin{cases} t_e(l) & (l = 1) \\ t_e(l) + \sum_{i=1}^{l-1} t_n(i) & (l > 1) \end{cases}, \quad (4)$$

where $t_e(l)$ represents the time required for the person to find the document in the l-th stack. On the other hand, $t_n(i)$ represents the time required for the person to realize that the document does not exist in the i-th stack.

Because there are generally multiple document stacks on a desktop (i.e., $L > 1$) as noted above, the person will find the desired document after searching through more than one stack in most cases (i.e., $l > 1$). Therefore, it is more significant to make $t_n(i)$ shorter than to make $t_e(l)$ shorter.

5.3.2 Method

We prepared a document stack x, and measured $t_n(x)$ and $t_e(x)$ through the user study. Each time was measured in the following three different document search techniques: Hand Search, Browse Search, and Thumbnail Search. Consequently, we measured six different document search time as shown in Table 2.

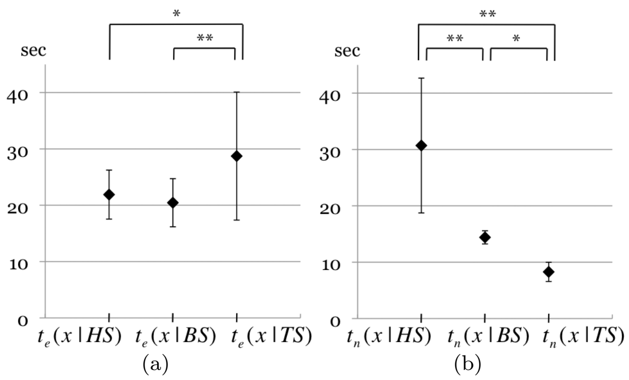


Fig. 12 Average document search time with standard deviation: (a) in the case where target documents existed and (b) in the case where they did not exist ($** : p < 0.01, * : p < 0.05$)

A document stack, a copy of a target document, and a timer were prepared in each trial. First, a participant was allowed to look at the copy and recognized that he (or she) would try to find this document in the trial. Then the participant started the document search. Simultaneously, the timer was activated. In cases where the target document did not exist in the stack, the participant stopped the timer when he (or she) realized it. When the document existed in the stack, the participant stopped the timer after finding and taking it. Note that the participant did not know whether the document existed in the stack at the beginning of each trial.

Six different document stacks were prepared for the six conditions. Each document stack consisted of twenty documents (ten paper documents, eight magazines, one bookbinder, and one file folder). It took 0.5 [s] to make each document transparent in Browse Search and Thumbnail Search. The target was a paper document inserted at the sixteenth layer of the stack when we measured $t_e(x|HS)$, $t_e(x|BS)$, and $t_e(x|TS)$. Sixteen participants were recruited from a local university. Each performed the trials under all six conditions. Before the study, they used the proposed system to know how to use it.

5.3.3 Result

Figure 12(a) and (b) show the average document search time in a case where target documents existed and in the other case where they did not exist, respectively. A one-way analysis of variance (ANOVA) with repeated measures showed statistically significant differences of all the respective times in Fig. 12(a) ($F_{2,30} = 6.9, p < 0.01$) and Fig. 12(b) ($F_{2,30} = 41.8, p < 0.01$). Post-hoc analysis was then performed using a Student-Newman-Keuls test for pairwise comparison. It showed statistically significant differences between each pair of conditions in both cases except for between $t_e(x|HS)$ and $t_e(x|BS)$ as shown in Fig. 12.

Fig. 12(a) shows that Browse Search did not significantly improve the document search time over Hand Search when the target document existed in the stack. Furthermore, Hand Search was about 7 [s] faster than Thumbnail Search. On the other hand, when the target document did not exist in the stack, the document search time could be significantly shortened by both Browse Search (about 16 [s]) and Thumbnail Search (about 22 [s]) over Hand Search (Fig. 12(b)).

In summary, Browse Search could shorten $t_n(i)$ over Hand Search. On the other hand, longer $t_e(l)$ was needed in Thumbnail Search than in Hand Search, although $t_n(i)$ could be significantly shortened in Thumbnail Search. The amount of time shortened in $t_n(i)$ (22 [s]) was more than thrice as great as the amount of time increased in $t_e(l)$ (7 [s]). Therefore, according to the proposed model in 5.3.1, both the proposed techniques can shorten a document search time when a person searches for a document from at least two document stacks. So, when $l = 1$,

$$T(l|L, HS) = T(l|L, BS) < T(l|L, TS), \quad (5)$$

and when $l > 1$,

$$T(l|L, HS) > T(l|L, BS) > T(l|L, TS), \quad (6)$$

where $T(l|L, HS)$, $T(l|L, BS)$, and $T(l|L, TS)$ represent the total amounts of time required in Hand Search, Browse Search, and Thumbnail Search, respectively. Because there are generally multiple document stacks on a desktop, we believe that the proposed techniques can effectively support document search.

5.4 Evaluation of Direct Select

We carried out another user study to evaluate the last interaction technique, Direct Select. The technique might be useful when a user is interested in a document in a stack and would like to check what it is. Without the technique, a person normally has to remove the upper documents by his (or her) hands only, even if they are quite heavy to move. We refer to this manipulation as **Hand Select**. This study was conducted to confirm whether or not Direct Select is more effective than Hand Select in such usage scenario.

5.4.1 Method

We compared the task completion time that was needed to recognize the month and year of publication of a magazine in a document stack by Direct Select, $t_r(DS)$, and by Hand Select, $t_r(HS)$. A document stack, a timer, a pen, and a sheet of paper were placed on a desktop. The document stack consisted of the same kinds of documents used in the previous study. The sixteenth document was a monthly magazine on which the month and year of publication were printed.

First, we told the participant to check the sixteenth document and to write down its month and year of publication. Then the participant started to do the task by either Direct Select or Hand Select. Simultaneously, the timer was activated. After the participant recognized the document and wrote the information on the paper, he (or she) stopped the timer.

Each participant performed two different trials: one was by Direct Select and the other was by Hand Select. We prepared a different document stack for each trial. Sixteen participants were recruited from a local university.

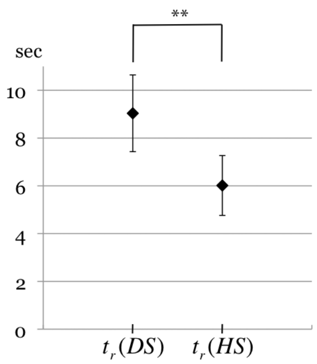


Fig. 13 Average task completion time with standard deviation in recognition of month and year of publication of magazine ($** : p < 0.01$)

5.4.2 Result

Figure 13 shows the average task completion time in each condition. A paired T test between them ($t_{15} = 13.98, p < 0.01$) showed that Direct Select could significantly shorten the time over Hand Select. Therefore, we reached the conclusion that Direct Select is more effective than Hand Select in a usage scenario where people would like to check a document in a document stack if its edge shows from the stack.

6 Discussion

The proposed touch sensing method can be applied to other projection-based MR systems where users interact with computer systems in the real world. A possible application is a projection-based real world painting system such as (Bandyopadhyay et al, 2001). In the previous study, a user must have an unusual input device to draw on a physical object. But the proposed method realizes that the user can paint physical objects directly by touching them. The system projects colored images onto the surface where touched areas are detected. Users can use their own hands, fingers, bodies, and even breath to paint on physical objects. They can also use a tool in their immediate environment, such as a paintbrush, an airbrush with hot water, and a hairdryer, as in (Iwai and Sato, 2005). Furthermore, the painting results can be unique because the hand shape or air flow appears on the physical object. This cannot be achieved by normal digital painting tools.

As used in (Kim et al, 2004), several vision-based methods can recognize and track documents on a desktop in real time based on feature point matching techniques (Lowe, 2004; Lepetit and Fua, 2006). By exploiting these methods, the proposed system can automatically recognize all documents in a stack even when documents are piled and removed dynamically on a desktop, as shown in (Kim et al, 2004).

We believe that the proposed framework can be used for applications other than document search. For example, the proposed system can be applied to an educational interface where students can look inside various living things (e.g., plants, animals, and human body) and machines (e.g., printers, PCs, TVs, audio players, air-conditioners, and cars) with an intuitive touch interface.

This paper presents a proof-of-concept system. Although the current configuration is large and expensive, all the devices it uses have been getting cheaper and smaller recently. Therefore, we believe that the system will be accepted in a typical office in the near future.

7 Conclusion

This paper has presented *Limpid Desk*, which supports document search on a physical desktop by making the upper layer of a document stack transparent in a projection-based MR environment. A user can visually access a lower-layer document without physically removing the upper documents. This is accomplished by superimposition of cover textures of lower-layer documents on the upper documents by projected imagery. This paper has introduced a method of generating projection images that make physical documents appear transparent. Furthermore, a touch sensing method based on thermal trace detection has been proposed for the system's input interface. User's touched areas on physical documents can be detected without any user-worn or hand-held devices. This interface allows a user to select a stack to be made transparent by a simple touching gesture. Three document search support techniques have been proposed and the results of a user study showed the effectiveness of the proposed technique.

In future, we will try to extend the proposed framework to seamlessly combine the virtual and physical desktop spaces.

References

Bandyopadhyay D, Raskar R, Fuchs H (2001) Dynamic Shader Lamps: Painting on Movable Objects. In: Proc. of IEEE/ACM Int'l Symp. on Augmented Reality (ISAR'01), pp 207–216

Baudisch P, Gutwin C (2004) Multiblending: Displaying Overlapping Windows Simultaneously without The Drawbacks of Alpha Blending. In: Proc. of ACM Conf. on Human Factors in Computing Systems (CHI '04), pp 367–374

Bimber O, Raskar R (2005) Spatial Augmented Reality: Merging Real and Virtual Worlds. A. K. Peters Ltd.

Bimber O, Iwai D, Wetzstein G, Grundhöfer A (2008) The Visual Computing of Projector-Camera Systems. Computer Graphics Forum 27(8):2219–2254

Bonanni L, Lee CH, Selker T (2005) Attention-Based Design of Augmented Reality Interfaces. In: Proc. of ACM Conf. on Human Factors in Computing Systems (CHI '05) Extended Abstracts, pp 1228–1231

Dietz P, Leigh D (2001) DiamondTouch: A Multi-User Touch Technology. In: Proc. of ACM Symp. on User Interface Software and Technology (UIST '01), pp 219–226

Grundhöfer A, Bimber O (2008) Real-Time Adaptive Radiometric Compensation. IEEE Transactions on Visualization and Computer Graphics 14(1):97–108

Ho HN, Amemiya T, Ando H (2007) Revealing Invisible Traces of Hand-Object Interactions with Thermal Vision. In: Proc. of the 2007 InfraMation Conference, pp 431–438

Inami M, Kawakami N, Tachi S (2003) Optical Camouflage Using Retro-reflective Projection Technology. In: Proc. of IEEE/ACM Int'l Symp. on Mixed and Augmented Reality (ISMAR '03), pp 348–349

Iwai D, Sato K (2005) Heat Sensation in Image Creation with Thermal Vision. In: Proc. of ACM Int'l Conf. on Advances in Computer Entertainment Technology (ACE '05), pp 213–216

Iwai D, Sato K (2006) Limpid Desk: See-Through Access to Disorderly Desktop in Projection-Based Mixed Reality. In: Proc. of ACM Symp. on Virtual Reality Software and Technology (VRST '06), pp 112–115

Iwai D, Hanatani S, Horii C, Sato K (2006) Limpid Desk: Transparentizing Documents on Real Desk in Projection-Based Mixed Reality. In: Proc. of IEEE Workshop on Emerging Display Technologies (EDT '06), pp 30–31

Kim J, Seitz SM, Agrawala M (2004) Video-Based Document Tracking: Unifying Your Physical and Electronic Desktops. In: Proc. of ACM Symp. on User Interface Software and Technology (UIST '04), pp 99–107

Koike H, Sato Y, Yoshinori K (2001) Integrating Paper and Digital Information on EnhancedDesk: A Method for Realtime Finger Tracking on an Augmented Desk System. ACM Transactions on Computer-Human Interaction 8(4):307–322

Lepetit V, Fua P (2006) Keypoint Recognition using Randomized Trees. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(9):1465–1479

Lowe DG (2004) Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 60(2):91–110

Sato K, Inokuchi S (1987) Range-Imaging System Utilizing Nematic Liquid Crystal Mask. In: Proc. of IEEE Int'l Conf. on Computer Vision (ICCV '87), pp 657–661

Seetzen H, Heidrich W, Stuerzlinger W, Ward G, Whitehead L, Trentacoste M, Ghosh A, Vorozcovs A (2004) High dynamic range display systems. In: Proc. of ACM Int'l Conf. on Computer Graphics and Interactive Techniques (SIGGRAPH '04), pp 760–768

Siio I, Rawan J, Mynatt E (2003) Finding Objects in "Strata Drawer". In: Proc. of ACM Conf. on Human Factors in Computing Systems (CHI '03) Extended Abstracts, pp 982–983

Swain MJ, Ballard DH (1991) Color Indexing. International Journal of Computer Vision 7(1):11–32

Wilson AD (2005) PlayAnywhere: A Compact Interactive Tabletop Projection-Vision System. In: Proc. of ACM Symp. on User Interface Software and Technology (UIST '05), pp 83–92

Yasuda K, Naemura T, Harashima H (2004) Thermo-Key: Human Region Segmentation from Video. IEEE Computer Graphics and Applications 24(1):26–30

Yoshida T, Horii C, Sato K (2003) A Virtual Color Reconstruction System for Real Heritage with Light Projection. In: Proc. of Int'l Conf. on Virtual Systems and Multimedia (VSMM '03), pp 161–168

Zollmann S, Bimber O (2007) Imperceptible Calibration for Radiometric Compensation. In: Proc. of Eurographics (short paper), pp 61–64

A Radiometric Compensation

In the real world, most objects have spatially varying reflectance properties that disturb the appearance of a projected image. We apply a radiometric compensation method (Yoshida et al, 2003), so as not to decrease the visibility of the appearance content of the projected thermal image. Note that the method assumes the spectral response of the visible camera is same as one of the human eye.

The method uses the visible camera of the system to obtain an affine correlation in color space between the projection and the reflectance captured by the camera for each camera pixel. This affine correlation can transform the desired color appearance on a physical object's surface into a projected color value. When the input RGB value for the projector is represented as (I_R, I_G, I_B) and the captured RGB value of the visible camera is represented as (C_R, C_G, C_B), the correlation between them can be represented by the following equation in the affine transformation.

$$\left[\, I_R \ I_G \ I_B \,\right]^t = \mathbf{K} \left[\, C_R \ C_G \ C_B \ 1 \,\right]^t . \tag{7}$$

$\mathbf{K}$ is a 3×4 matrix that transforms a camera's color space to that of a projector. Therefore, $\mathbf{K}$ is called a color mixing matrix that takes into account the projector's spectral characteristics, the camera's spectral sensitivity, and the spectral reflectance of the object's surface. $\mathbf{K}$ has to be calibrated for each camera pixel. Once at least four correspondences between (C_R, C_G, C_B) and (I_R, I_G, I_B) are obtained, $\mathbf{K}$ is calculated by a least-squares method. In the calibration process, more than four simple color patterns (e.g., red, green, blue, yellow, magenta, and cyan) are projected and the reflectance of each projected pattern is captured. After this color calibration, images of the desired color can be displayed on surfaces under consideration of their reflectance. The color of the compensated projection image (I_R, I_G, I_B) is calculated by Eq. (7) for each pixel of the projection image where RGB values of the desired color are assigned to (C_R, C_G, C_B). As described above, the color mixing matrix $\mathbf{K}$ can be calibrated without any prior information about the spectral characteristics of the projector, the camera, and the object's surface.