AR IN VR: SIMULATING INFRARED AUGMENTED VISION

Fayez Lahoud   *Sabine Süsstrunk*

School of Computer and Communication Sciences, EPFL, Switzerland

ABSTRACT

Developing an augmented reality (AR) system involves multiple algorithms such as image fusion, camera synchronization and calibration, and brightness control, each of them having diverse parameters. This abundance of settings, while allowing for many features, is detrimental to developers as they try to navigate between different combinations and pick the most suitable towards their application. Additionally, the temporally inconsistent nature of the real world makes it hard to build reproducible scenarios for testing and comparison. To help address these issues, we develop a virtual reality (VR) environment that allows simulating a variety of AR configurations[1]. We show the advantages of AR simulation in virtual reality, demonstrate an image fusion AR system and conduct an experiment to compare different fusion methods.

Index Terms— augmented reality, virtual reality, augmented vision, image fusion, simulation, infrared

1. INTRODUCTION

The recent growth in augmented reality (AR) motivated a wide range of wearable headsets such as Microsoft HoloLens and Epson Moverio. They have been widely used in various applications in industry and entertainment. One subset of AR applications that has recently gained interest is vision augmentation through optical see-through head mounted displays (OST HMD). It has already been used for numerous vision enhancements, such as vision correction [1] and night vision [2]. One particular application of image fusion in AR is thermal imaging [3, 4], already employed for firefighting [5, 6], military, and police organizations.

However, the majority of commercially available infrared (IR) cameras have lower frame rate and resolution than the typical color ones, thus creating spatial and temporal registration inconsistencies. This creates manual interaction difficulties, occlusion of details between thermal and color images, false colors, and possibly user disorientation. Hence, it is imperative to choose a suitable fusion method to alleviate these issues. Previous studies on coloration and fusion techniques [7, 8] were limited to static images and were not designed to operate in dynamic lighting conditions or in an AR setting.

In the case of firefighting, it is of critical importance to study these factors and classify their effects on task performance. However, it is challenging to conduct controlled comparative studies due to reproducibility issues that arise from the ever-changing real world background and the danger of using untested methods in a fire situation.

In this paper, we propose using an AR simulation of image fusion in virtual reality (VR) to circumvent these difficulties. We use Unity3D [9] and Oculus Rift [10] to build the simulation. The novelty in our approach is the simulation of a thermal camera as well as a configurable AR display. The AR display lies on top of the user's view of the virtual world replicating an OST HMD. Furthermore, we implement several existing fusion methods and conduct a comparative experiment to test their performance in a search and rescue task using the developed simulation.

2. RELATED WORK

Our main goal is facilitating AR testing for firefighters. We study how AR simulation can be used to drive the development of such AR systems.

SidebARs [11] provides distance and direction to reach important locations in the scene. In [12], the authors display fire locations and building plans in AR. Both works report the difficulty of designing a system aimed at firefighters. Our proposed method simplifies the presentation and preview of changes in the system and improves the development process.

In [13,14], the authors research image fusion methods and their effect on performance. However, each method is singled out and tested without context. Additionally, the images used for testing were static and do not represent the dynamic world encountered in AR applications. Our simulation scales and generalizes better with the number of tested settings and explores the fusion algorithms under a dynamic environment with varying lighting conditions.

VisMerge [15] proposes an AR simulation to compare fusion methods. They rely on the 3D objects vertices to compute the fusion, while ours simulates a thermal camera instead and reproduces realistic infrared imagery. Consequently, we are able to test fusion between 2D infrared and color images. Additionally, they display the fusion on top of the whole user's field of view while we simulate a configurable AR display.
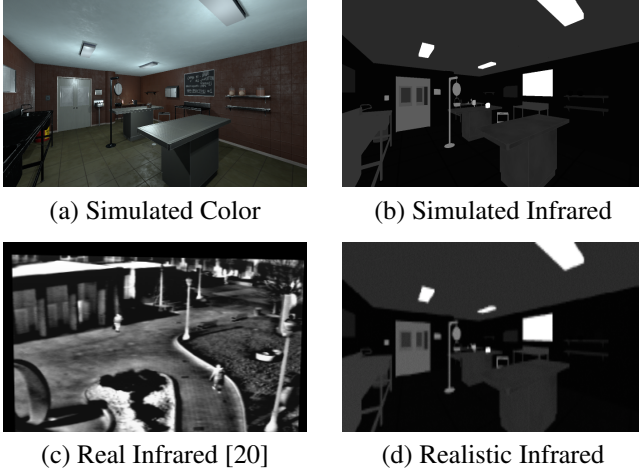
---

[1]Code can be found at https://infoscience.epfl.ch/record/255390/files/

(a) Simulated Color   (b) Simulated Infrared

(c) Real Infrared [20]   (d) Realistic Infrared

Fig. 1. Real and simulated color and infrared images

Fig. 2. Augmented vision simulation

3. BENEFITS OF AR IN VR

Lee et al. [16] replicated a previously held AR experiment in a VR simulation and showed that AR in VR can reproduce reliable and authentic results. Consequently, AR simulation can solve the difficulties encountered in developing the AR system. In fact, it has been previously adopted and successfully used to study the effect of background illumination over legibility [17], the performance of a car's AR windshield [18] and the effects of field of view in X-Ray vision [19].

First, the user has complete control over the environment as opposed to the real world. Experiments can be repeatedly tested with the same parameters. Furthermore, properties such as display position and field of view can be modeled and compared. Testing system designs becomes free of expensive low level implementations and additional devices. Finally, the developers can simulate a multitude of scenarios, including those that might be difficult, dangerous, or costly to reproduce in a fire situation.

4. SIMULATED THERMAL CAMERA

In Unity [9], shaders take material properties and compute how they should be displayed on the graphical processing unit (GPU). A replacement shader uses the same properties but runs a different computation. We will use the normal shader to render the color image and a replacement for the thermal image.

First, we assign temperature and emission properties to all the objects in the scene. The emission property helps differentiate materials at the same temperature, e.g. wood and metal won't behave similarly even at the same temperature. Then, a camera replacement shader looks at these specific values for each object and renders an infrared map of the scene, such as Fig.1b, while the normal shader renders Fig.1a.

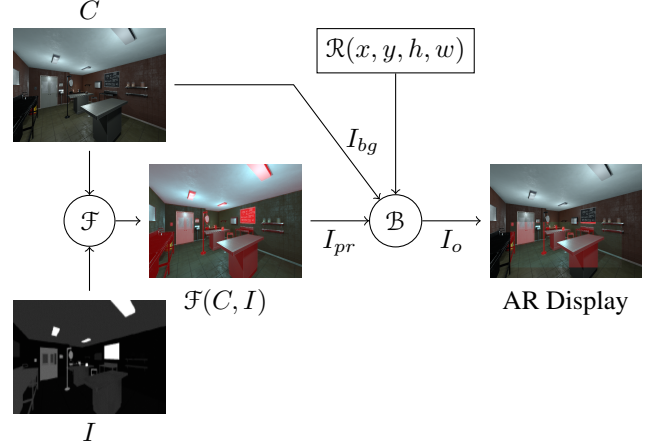In thermal imagery such as Fig.1c, energy is emitted from

the surface of objects and heat moves through solids, liquids and gazes removing sharp edges. Additionally, the thermal imaging camera (TIC) is constantly moving in the scene giving blurry captures. Furthermore, TICs exhibit noisy images due to multiple factors such as their thermal range, the ambient temperature and the detector sensitivity. Finally, current TIC technologies in the market have a relatively small resolution due to cost and large pixel detectors. Most common resolutions are 160x120 and 320x240. Applying blur, noise and low resolution to the output gives us more realistic simulations similar to Fig.1d.

5. AUGMENTED REALITY SIMULATION

Image fusion is the process of combining information from two or more input images into a single one. The goal is to display more information in one image than any of the single input images. Additionally, we only want to show the information on a subsection of the screen to simulate an AR display.

Using our developed setup, we simulate both color C and infrared I images of the scene. An algorithm $\mathcal{F}$ computes their fusion according to Fig.2. We limit the AR display to a rectangle $\mathcal{R}(x, y, h, w)$ inside the screen that acts as a combiner to additively blend the projector light I_{pr} resulting from $\mathcal{F}(C, I)$ with the background light I_{bg}. The output I_o for every pixel i is computed as follows:

$$I_o^i = \begin{cases} r * I_{pr}^i + t * I_{bg}^i & \text{if } i \in \mathcal{R}(x, y, w, h) \\ I_{bg}^i & \text{otherwise} \end{cases} \quad (1)$$

where r and t are the reflectivity and transmissivity properties of the combiner. We set $r = 0.95$ and $t = 0.46$ according to [21].

For our simulation, we implement six fusion methods from [15] shown in Fig.3. **Spectral Compression** $\mathcal{F}_1$ compresses the thermal and color information into the three RGB
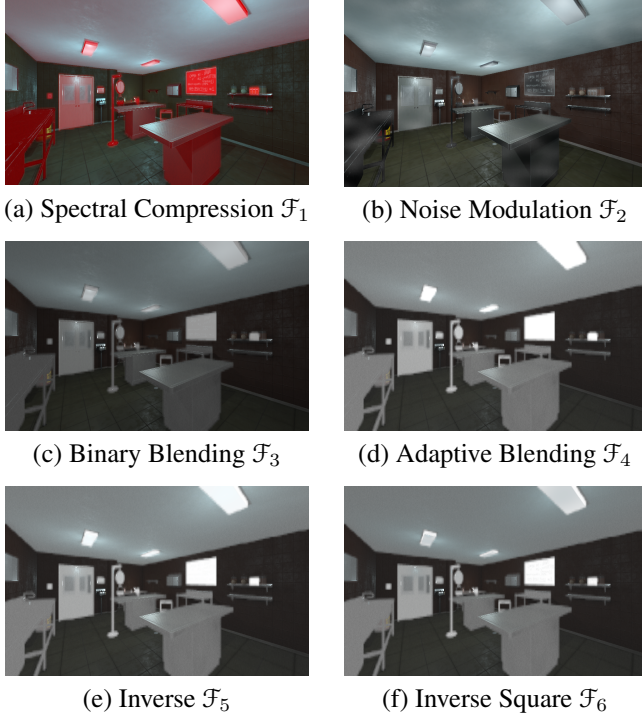
(a) Spectral Compression $\mathcal{F}_1$ (b) Noise Modulation $\mathcal{F}_2$

(c) Binary Blending $\mathcal{F}_3$ (d) Adaptive Blending $\mathcal{F}_4$

(e) Inverse $\mathcal{F}_5$ (f) Inverse Square $\mathcal{F}_6$

Fig. 3. The different tested fusion algorithms

Scenario	Run Name	Fusion	Light	Smoke
Fusion	F-Spectral	$\mathcal{F}_1$	Normal	Off
	F-Noise	$\mathcal{F}_2$		
	F-Binary	$\mathcal{F}_3$		
	F-Adaptive	$\mathcal{F}_4$		
	F-Inverse	$\mathcal{F}_5$		
	F-InverseSq	$\mathcal{F}_6$		
Light	L-Normal	$\mathcal{F}_{best}$	Normal	Off
	L-Dark		Dark	
	L-Bright		Bright	
Smoke	S-Normal	$\mathcal{F}_{best}$	Normal	Off
	S-Cold			Cold
	S-Hot			Hot

Table 1. Scenarios summary

Users were asked to participate in three scenarios listed in Table 1. Each one consists of multiple 30 second runs with varying configurations. In the first, we only vary the image fusion algorithm under normal visibility. In the latter two, the participant is exposed to variations of light and smoke, respectively. Additionally, a first training scenario is conducted to acclimate the participants with the virtual environment and help them understand how to navigate and locate targets. Different conditions can be viewed in Fig.4. Finally, all the scenarios were run on the same day, and users were exposed to the runs within a scenario in a random order, each run once.
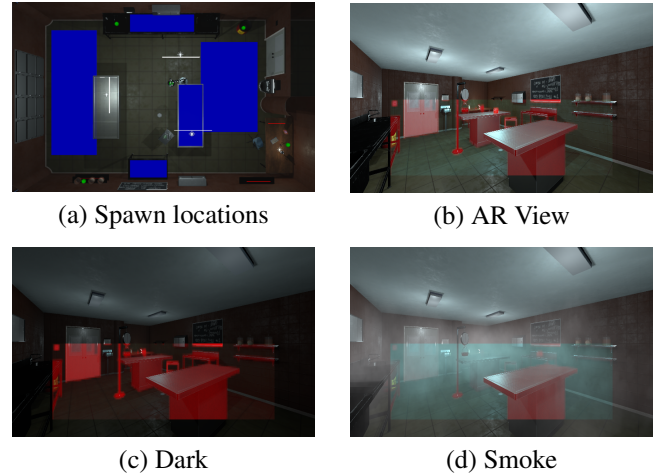
channels creating a false red-cyan coloring scheme for hot-cold temperatures. **Noise Modulation** $\mathcal{F}_2$ simply adds white noise to the thermally hot regions of the image, thus the warmer and colder areas can be identified by their noisy component. **Binary Blending** $\mathcal{F}_3$ blends both thermal and color images using a predefined blending factor while **Adaptive Blending** $\mathcal{F}_4$ calculates a per-pixel factor based on the contribution of color and thermal values. In **Inverse** $\mathcal{F}_5$ and **Inverse Square** $\mathcal{F}_6$, the blending is inversely proportional to the average pixel intensity, the latter intensifies this effect at extreme values.

6. EXPERIMENT

We conduct an experiment to compare performances for a target rescue task in low visibility conditions using different fusion methods.

6.1. Setup

We modified a pre-built room to accommodate our experiment, as well as a chest with a warm temperature as rescue target. All the other objects are stationary in the room and assigned a temperature according to their nature (e.g. lights are hot). The target's spawn location is randomly sampled according to Fig.4. The targets are split into 4 difficulty levels according to their visibility ranging from fully visible to completely occluded.

(a) Spawn locations   (b) AR View

(c) Dark   (d) Smoke

Fig. 4. Experimental Setup

6.2. Results

For the evaluation, 9 participants (between 20 and 30 years old) conducted the experiment. Out of the 9, 1 wasn't able to complete it due to nausea.
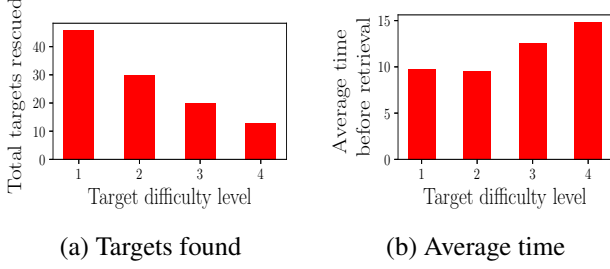
(a) Targets found   (b) Average time

Fig. 5. Total targets rescued and average retrieval time with respect to difficulty levels

(a) Fusion scenario   (b) Light and smoke scenarios

Fig. 6. Target scores for all scenarios

6.2.1. Weighting Scheme

The targets are split into 4 difficulty levels, so we first check the reasonableness of our classification. Fig.5 shows the total targets found and average time taken to find a target per level. These results validate the tiers distribution.

Using the previously defined difficulty classes, a weighting scheme scores targets by the effort needed to find them. Starting with 1 point for the lowest level, up to 4 points for the hardest tier. The time taken to find a target is also factored in the computation. In summary, the final score is computed according to:

$$S_i = max(0, D_i * (1 - \frac{T_i - T_{D_i}}{T_{D_i}})) \quad (2)$$

where D_i indicates the difficulty of target i, T_i the time taken to rescue it and T_{D_i} the average time for difficulty D_i.

6.2.2. Image Fusion Scenario

Using the weighted scores, we conduct an ANOVA test for the image fusion scenario. We obtain $F = 21.29$ and $p << 0.0.1$ showing that the image fusion algorithm did in fact affect the performance (score). Contrarily, we get $F = 2.56$ and $p = 0.02$ for varying the users. Thus, participants seem not to affect the results too much. Fig.6a shows the average rescue rate per fusion algorithm. **Noise Modulation**, **Inverse Square** and **Adaptive Blending** outperform the other three methods. There is an improvement of 33% moving from the worst method **Inverse** to the best **Noise Modulation**.

Comparing fusion methods, we see that by squaring the values in the **Inverse Square** method, we gain a performance boost of 30% over the **Inverse**. Pushing the contrast to extreme values proved helpful in directing the participants' attention. Similarly, there is an improvement of 23% between **Adaptive Blending** and **Binary Blending**, validating the need of adapting the blending factor according to the contributions from the RGB and IR channels.

6.2.3. Light and Smoke Scenarios

ANOVA tests for the light and smoke scenarios result in $F_l = 18.06$ and $F_s = 16.35$, respectively, with $p << 0.01$ for
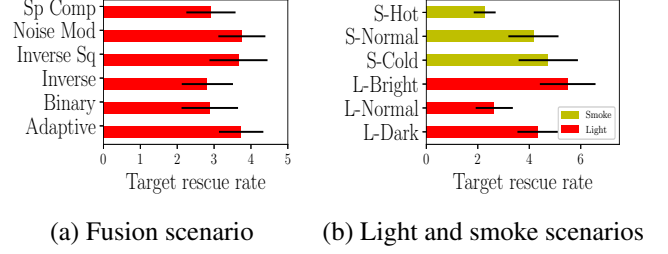
both. Thus, the light and smoke conditions affect participants' performance. Fig.6b shows the average weighted scores for all repetitions of these scenarios.

The participants made better use of the AR display in darker surround because their vision was limited and the contrast was stronger with the surroundings compared to normal lighting conditions. In bright light, the AR display was washed out due to the background brightness, however, users were still able to obtain high scores because the targets were very easy to spot. In the normal test case, users didn't perform exceptionally well, this could be because they didn't use the AR display as efficiently as they did in the darker setting.

Cold smoke slightly improved performance, the fusion algorithms aren't affected by regions with high levels of low temperature smoke. Contrarily, participants reported the worst scores in the hot smoke scenario, thus the used fusion methods weren't suitable for a hot smoke scenario.

Finally, participants encounter the same situation in **L-Normal** and **S-Normal** as they have the same light and smoke conditions. There is an improvement greater than 60% between those two runs. This indicates that participants have adapted to the situation, and understood how to use the image fusion to their benefit. Consequently, the simulation itself could potentially be used to train users by repetition.

7. CONCLUSION

In this paper, we present an AR simulation in VR. We first simulate a thermal camera and an AR display to show the fusion results. We also conduct a comparative experiment to evaluate the performance of different fusion methods under varying light and smoke conditions.

In the future, we will improve the AR display and the color and thermal cameras by more accurately simulating physical light properties. We will also expand out experiments to cover a larger set of variables and scenarios.

8. ACKNOWLEDGMENTS

9. REFERENCES

[1] Eli Peli, Gang Luo, Alex Bowers, and Noa Rensing, "Applications of augmented-vision head-mounted systems in vision rehabilitation," *Journal of the Society for Information Display*, vol. 15, no. 12, pp. 1037–1045, 2007.

[2] Chunjia Hu, Guangtao Zhai, and Duo Li, "An augmented-reality night vision enhancement application for see-through glasses," in *Multimedia & Expo Workshops (ICMEW), 2015 IEEE International Conference on*. IEEE, 2015, pp. 1–6.

[3] Jungong Han, Eric J Pauwels, and Paul De Zeeuw, "Visible and infrared image registration in man-made environments employing hybrid visual features," *Pattern Recognition Letters*, vol. 34, no. 1, pp. 42–51, 2013.

[4] Fei Liu and Stefan Seipel, "Infrared-visible image registration for augmented reality-based thermographic building diagnostics," *Visualization in Engineering*, vol. 3, no. 1, pp. 16, 2015.

[5] Steven K Feiner, "Augmented reality: A new way of seeing," *Scientific American*, vol. 286, no. 4, pp. 48–55, 2002.

[6] Mathias Kolsch, Ryan Bane, Tobias Hollerer, and Matthew Turk, "Multimodal interaction with a wearable augmented reality system," *IEEE Computer Graphics and Applications*, vol. 26, no. 3, pp. 62–71, 2006.

[7] R Istenic, D Heric, S Ribaric, and Damjan Zazula, "Thermal and visual image registration in hough parameter space," in *Systems, Signals and Image Processing, 2007 and 6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services. 14th International Workshop on*. IEEE, 2007, pp. 106–109.

[8] Saeed Yahyanejad and Bernhard Rinner, "A fast and mobile system for registration of low-altitude visual and thermal aerial images using multiple small-scale uavs," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 104, pp. 189–202, 2015.

[9] 2017 Unity Technologies, "Unity 3D," .

[10] 2017 Oculus VR, LLC, "Oculus rift," .

[11] Teresa Siu and Valeria Herskovic, "Sidebars: improving awareness of off-screen elements in mobile augmented reality," in *Proceedings of the 2013 Chilean Conference on Human-Computer Interaction*. ACM, 2013, pp. 36–41.

[12] Daniel Steingart, Joel Wilson, Andrew Redfern, Paul Wright, Russell Romero, and Lloyd Lim, "Augmented cognition for fire emergency response: An iterative user study," in *1st International Conference on Augmented Cognition*, 2005.

[13] Harishwaran Hariharan, Andrei Gribok, Mongi A Abidi, and Andreas Koschan, "Image fusion and enhancement via empirical mode decomposition," *Journal of Pattern Recognition Research*, vol. 1, no. 1, pp. 16–32, 2006.

[14] David Connah, Mark Samuel Drew, and Graham David Finlayson, "Spectral edge image fusion: Theory and applications," in *European Conference on Computer Vision*. Springer, 2014, pp. 65–80.

[15] Jason Orlosky, Peter Kim, Kiyoshi Kiyokawa, Tomohiro Mashita, Photchara Ratsamee, Yuki Uranishi, and Haruo Takemura, "Vismerge: Light adaptive vision augmentation via spectral and temporal fusion of non-visible light," in *Mixed and Augmented Reality (ISMAR), 2017 IEEE International Symposium on*. IEEE, 2017, pp. 22–31.

[16] Cha Lee, Scott Bonebrake, Tobias Hollerer, and Doug A Bowman, "A replication study testing the validity of ar simulation in vr for controlled experiments," in *Mixed and Augmented Reality, 2009. ISMAR 2009. 8th IEEE International Symposium on*. IEEE, 2009, pp. 203–204.

[17] Joseph L Gabbard, II J Edward Swan, and Deborah Hix, "The effects of text drawing styles, background textures, and natural lighting on text legibility in outdoor augmented reality," *Presence: Teleoperators and Virtual Environments*, vol. 15, no. 1, pp. 16–32, 2006.

[18] SeungJun Kim and Anind K Dey, "Simulated augmented reality windshield display as a cognitive mapping aid for elder driver navigation," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2009, pp. 133–142.

[19] Jonathan Ventura, Marcus Jang, Tyler Crain, Tobias Höllerer, and Doug Bowman, "Evaluating the effects of tracker reliability and field of view on a target following task in augmented reality," in *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*. ACM, 2009, pp. 151–154.

[20] James W Davis and Vinay Sharma, "Background-subtraction using contour-based fusion of thermal and visible imagery," *Computer Vision and Image Understanding*, vol. 106, no. 2, pp. 162–182, 2007.

[21] Jason Rmy Racine, "Evaluation of user experiences with wearable augmented reality in fire fighting operations," Ecole Polytechnique Fdrale de Lausanne, I&C school, 2016.