Multispectral Invisible Coating: Laminated Visible-Thermal Physical Attack against Multispectral Object Detectors Using Transparent Low-E Films

Taeheon Kim, Youngjoon Yu, Yong Man Ro*

Image and Video Systems Lab, KAIST, South Korea
{eetaekim, greatday, ymro}@kaist.ac.kr

Abstract

Multispectral object detection plays a vital role in safety-critical vision systems that require an around-the-clock operation and encounter dynamic real-world situations (e.g., self-driving cars and autonomous surveillance systems). Despite its crucial competence in safety-related applications, its security against physical attacks is severely understudied. We investigate the vulnerability of multispectral detectors against physical attacks by proposing a new physical method: Multispectral Invisible Coating. Utilizing transparent Low-e films, we realize a laminated visible-thermal physical attack by attaching Low-e films over a visible attack printing. Moreover, we apply our physical method to manufacture a Multispectral Invisible Suit that hides persons from the multiple view angles of Multispectral detectors. To simulate our attack under various surveillance scenes, we constructed a large-scale multispectral pedestrian dataset which we release to the public. Extensive experiments show that our proposed method effectively attacks the state-of-the-art multispectral detector both in the digital space and the physical world.

Introduction

Multispectral object detection plays an important role in safety-critical applications such as autonomous surveillance systems and self-driving cars (Farlik et al. 2019; Choi et al. 2018; Zhang et al. 2019; Cao et al. 2021). The motivation behind multispectral object detection is combining information obtained from multispectral cameras (e.g., visible and thermal) to handle dynamic situations in the real world, especially low visibility conditions such as bad weather, low illumination, and low resolution (Li et al. 2019; Zhou, Chen, and Cao 2020). Combining thermal imaging, which leverages thermal heat emission to capture the silhouette of objects, is the most intuitive way of perceiving objects under low visibility conditions (Krišto, Ivasic-Kos, and Pobar 2020; Kieu et al. 2020). Based on this intuition, researchers recently developed DNN-based multispectral object detectors that adequately conjugate the unique visible and thermal characteristics. By switching the input port between visible or thermal mode interchangeably depending on the situation, multispectral detectors show significant performance superiority against visible-only or thermal-only detectors (Wu et al. 2020a; Kim, Park, and Ro 2021, 2022; Park et al. 2021). Recent works show the remarkable success of multispectral detectors in practical day-night vision applications that encounter challenging real-world scenes such as pedestrian detection in driving or surveillance scenes.

Along with the importance and remarkable achievement in multispectral object detection, DNN-based object detectors are shown to be vulnerable to adversarial patch attacks (Liu et al. 2018; Chen et al. 2018; Thys, Van Ranst, and Goedemé 2019; Lee and Kolter 2019; Wang et al. 2021; Xu et al. 2020; Wu et al. 2020b; Kim, Yu, and Ro 2022; Yu et al. 2022). Adversarial patches are localized perturbations intentionally crafted by malicious attackers that fool machine learning models and lead to misprediction. It can be physically realized as stickers, printings, laser beams, or even wearable clothing (Li, Schmidt, and Kolter 2019; Thys, Van Ranst, and Goedemé 2019; Duan et al. 2021; Wu et al. 2020b). During the COVID-19 pandemic, thermal detectors received significant attention, and thermal adversarial patches were developed using infrared stealth materials or temperature controls (Zhu et al. 2021, 2022). Kim, Lee, and Ro (2022) designed a Multispectral Adversarial Patch by combining visible and thermal patch attacks. Adversarial patches are developing in various forms, imposing a notable threat to real-world detectors. For instance, a bank
robber can wear adversarial clothing to hide from an autonomous surveillance system, or a villain can invade the vision systems of self-driving cars by shooting adversarial laser beams. To prevent such potentially disastrous consequences caused by adversarial patches, investigating physical attacks and developing robust models are necessary for real-world detectors.

In this paper, we propose a new physical attack method against multispectral detectors. With multispectral detectors, if one source is attacked, the unattacked source can still be utilized for correct predictions; therefore, attacking all sources (visible & thermal) is necessary. Kim, Lee, and Ro (2022) achieved this by placing visible and thermal adversarial patches side-by-side. But that method has obvious limitations. The overall size of the patch is doubled, and metallic materials composing the thermal patch make the patch large and cumbersome. Our goal is to generate a lightweight single piece of flexible coating within a compact design that attacks visible and thermal sources simultaneously. To achieve our goal, our core idea is to utilize a new material: Transparent Low-e films. Originally, transparent Low-e films are window films manufactured to obtain heat-insulating properties for solar control in homes while preserving the view outside the window. Based on its original functionality, we exploit two excellent physical properties of Transparent Low-e films for our purpose. First, transparent Low-e films are thermal-insulating materials useful for representing different levels of thermal intensities, which are necessary to achieve strong thermal attacks. Second, transparency (visible light transmittance) is high (> 70%) such that the majority of the visible information behind the film can be transferred through the Low-e film.

Based on the properties above, we physically implement the visible-thermal attack by laminating Low-e films over the printed visible attack pattern. Furthermore, Low-e films are a self-adhesive and flexible material that can be laminated on any object; we apply our physical method to manufacture a Multispectral Invisible Suit. Figure 1 shows a person evading a multispectral detector by wearing a Multispectral Invisible Suit. Also, to evaluate our physical attack in surveillance environments, we construct a large-scale multispectral pedestrian surveillance dataset, which we will publicly release. It contains 3000 visible/thermal day/night pedestrian image pairs with corresponding manual annotations. We test our proposed attack on the FLIR ADAS dataset (FLIR Systems 2021) for driving scenes, and our collected dataset for surveillance scenes. Extensive experimental results show that multispectral invisible coating effectively hides objects from the multispectral detector both in digital space and the physical world. We expect our open dataset and the presented experiments will provide a benchmark for future research on developing robust multispectral detectors. The following summarizes our contributions:

- We propose a new physical attack method, “Multispectral Invisible Coating” a laminated visible-thermal attack leveraging a new material: Transparent Low-e films.
- Applying the Multispectral Invisible Coating onto clothing, we manufactured a Multispectral Invisible Suit that hides persons from multiple view angles of Multispectral detectors in the physical world.
- We constructed a publicly available large-scale multispectral pedestrian dataset that contains 3000 visible/thermal day/night image pairs captured from various surveillance scenes.

Related Work

Multispectral Object Detection

Recently, multispectral detectors have shown remarkable advantages, especially for all-day vision systems (Park, Kim, and Sohn 2018; Zhou, Chen, and Cao 2020; Guan et al. 2018; Marnissi et al. 2022; Liu et al. 2021). The release of multispectral pedestrian datasets (Hwang et al. 2015; González et al. 2016) motivated the computer vision community to advance the state-of-the-art vision models by additionally utilizing thermal input data to compensate for limitations of vision systems based on visible perception. Research on multispectral detection has actively progressed to adequately associate these two modalities to encode richer feature representations of objects. Liu et al. (2016) designed four distinct fusion architectures that fuse visible and thermal features on different branches of the Faster R-CNN network. Illumination-aware Faster R-CNN (Li et al. 2019) adaptively aggregates visible and thermal sub-networks to produce final prediction scores by a gate function that leverages the illumination value. Zhang et al. (2019) introduced the misclassification problem in multispectral detection tasks due to different Field-of-view (FOV) and frame rates between visible and thermal camera sensors. Kim, Park, and Ro (2021) mitigated this problem by designing an uncertainty-aware network based on Faster R-CNN. To the best of our knowledge, this model (Kim, Park, and Ro 2021) is currently the state-of-the-art multispectral pedestrian detector that we will select as our target model.

Adversarial Patch Attacks

Adversarial patch attacks are localized perturbations capable of fooling the prediction of DNN-based models which can be physically realized. In this section, we introduce adversarial patches on visible and thermal detectors, respectively. On visible detectors, Thys, Van Ranst, and Goedemé (2019) introduced an adversarial patch that can be printed on paper with a laser printer. Considering the deformation of clothes by Thin Plate Spline (TPS) transformation, adversarial t-shirts (Xu et al. 2020) and cloaks (Wu et al. 2020b) were developed which successfully fooled detectors. Recent works include realistic and natural-looking adversarial patches that fool both human eyes and detection models (Hu et al. 2021; Tan et al. 2021). Compared to the abundance of research on visible detectors, studies on adversarial patches on thermal detectors are yet limited. Zhu et al. (2021) designed a thermal adversarial patch by arranging small bulbs on a board. Zhu et al. (2022) manufactured an Infrared Invisible Clothing by attaching aerogel to clothing. Kim, Lee, and Ro (2022) crafted a Multispectral Adversarial Patch by placing a thermal adversarial patch consisting of aluminum, steel, and sandpaper alongside a printed visible patch.
Method

Attack Design
We design the Multispectral Invisible Coating as \( \hat{c}_{vis} \in \mathbb{R}^{H' \times W' \times C'} \) and \( \hat{c}_{th} \in \mathbb{R}^{H' \times W' \times C'} \). \( \hat{c}_{vis} \) and \( \hat{c}_{th} \) is attached to objects in clean visible-thermal image pair \( x_{vis} \in \mathbb{R}^{H \times W \times C} \) and \( x_{th} \in \mathbb{R}^{H \times W \times C} \), producing perturbed image pair \( \hat{x}_{vis} \in \mathbb{R}^{H \times W \times C} \) and \( \hat{x}_{th} \in \mathbb{R}^{H \times W \times C} \). To simulate this process, we specify the locations of \( \hat{c}_{vis} \) and \( \hat{c}_{th} \) to be attached within \( \hat{x}_{vis} \) and \( \hat{x}_{th} \) by a binary mask \( M \in \{0, 1\}^{H \times W \times C} \). The same binary mask is \( (M = M_{vis} = M_{th}) \) applied to \( \hat{c}_{vis} \) and \( \hat{c}_{th} \). Also, a transformation function \( A(c, t, x) \in \mathbb{R}^{H \times W \times C} \) is applied to simulate real-world deformations that occurs when the coating is attached to real objects in the physical world. Transforms \( t \in T \) include random cropping, Expectation of Transformations(EOT) and Thin Plate Spline(TPS) transforms. Our objective is to optimize \( \hat{c}_{vis} \) and \( \hat{c}_{th} \) such that the Multispectral detector \( D_{MS} \) cannot detect objects in \( \hat{x}_{vis} \) and \( \hat{x}_{th} \). The attack procedure of the Multispectral Invisible Coating can be stated as the following:

\[
\hat{x}_{vis} = (1 - M) \odot x_{vis} + M \odot A(\hat{c}_{th}, t, x_{th}) \\
\hat{x}_{th} = (1 - M) \odot x_{th} + M \odot A(\hat{c}_{vis}, t, x_{vis}) \\
\min_{x \sim X, t \sim T} \{D_{MS}(\hat{x}_{vis}, \hat{x}_{th})\}
\]

Utilizing Low-e Films for Thermal Attacks
To implement the physical thermal attack, we exploit the thermal radiation equation from a classical physics theory. According to the Stefan-Boltzmann law (Tiihonen 1997; Wellons 2007), the quantity of thermal radiation depends on the emissivity (\( \varepsilon \)) of the material and the surface quality under the same room temperature (\( T_{room} \)). Generally, the human body or fabric for clothing has high emissivity, whereas polished metals such as aluminum, copper, and silver have low emissivity. Using these properties, thermal attacks have been realized by arranging low emissivity materials to the human body (Kim, Lee, and Ro 2022; Zhu et al. 2022). Instead, we use special material, Low-e films, for our visible-thermal laminated attack. The Low-e film consists of a polyester film substrate that has micro-thin, transparent metal sputtering layers therefore it has high transparency in addition to low emissivity. Leading window film manufacturers such as Enerlogic® (Winckler 2012) and PEN-JEREX developed novel techniques and manufactured high-end products that are flexible, transparent, and have low emissivity (\( \varepsilon \)) around 0.05. We use Enerlogic®70 (\( \varepsilon = 0.06 \)) and a conventional Low-e film, 70RNE (\( \varepsilon = 0.4 \)) for our thermal attack with its properties shown in Table 1. Using these two types of films, we implement a 3-level intensity thermal attack, including the paper (or fabric) on which the visible attack is printed. Figure 2 illustrates our laminated attack using Low-e films.

Blending Low-e Films Over the Visible Printing
Our adversarial perturbation is computed digitally and then applied in the physical world. Therefore, during simulation, we need to consider the effect of laminating the Low-e films over the visible printing. Although Low-e films have high transparency, it has metallic color due to their polymer layers containing metals and metal oxides. Thus, the pixel values of the visible printing slightly change when the Low-e film is laminated over. The approximation of this effect can be achieved by an alpha blending process to represent \( \hat{c}_{vis} \), the visible perception of the Multispectral Invisible Coating from the visible camera. We alpha-blend between the RGB color of the Low-e film denoted as \( \hat{p}_{Lowe} \) and the visible printing \( \hat{p}_{vis} \). During this process, each pixel of \( \hat{p}_{vis} \) and \( \hat{p}_{Lowe} \) has an additional numeric value stored in its alpha channel (\( \alpha(i, j) \)). This value represents ranges from 0 to 1 representing how opaque each pixel is.

**Table 1: Physical properties of the Low-E film products used in our thermal attack.**

<table>
<thead>
<tr>
<th>Product</th>
<th>Emissivity</th>
<th>Transparency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enerlogic®70</td>
<td>0.06</td>
<td>70%</td>
</tr>
<tr>
<td>70RNE</td>
<td>0.4</td>
<td>69%</td>
</tr>
</tbody>
</table>

**Figure 2:** A piece of Multispectral Invisible Coating. Transparent Low-e films are attached over the visible printing. (a),(c) are attack patterns generated in the digital space. (b),(d) are physically manufactured multispectral invisible coating captured by a visible and thermal camera.
using $M_{Lowe}$ to obtain $\hat{c}_{vis}$ is defined as follows.

$$
\hat{c}_{vis}(i,j) = (\hat{p}_{vis}(i,j) \odot M_{Lowe}(i,j)) \ast (1 - \alpha(i,j))
+ (\hat{p}_{Lowe}(i,j) \odot M_{Lowe}(i,j)) \ast \alpha(i,j)
+ \hat{p}_{vis}(i,j) \odot (1 - M_{Lowe}(i,j))
$$

(5)

Figure 2 (a) shows digitally simulated $\hat{c}_{vis}(i,j)$ with $\alpha = 0.75$.

Simulating the Multispectral Invisible Coating in the Digital Space

The simulation process to generate Multispectral Invisible Coating is shown in Figure 3. To start with, we consider the real-world design process in making a suit using Multispectral Invisible Coating. We requested the tailor to manufacture the suit by periodically expanding the basic pattern. We simulate this process by enumerating the basic pattern $\hat{c}_{vis}$ and $\hat{c}_{th}$. We define the enumerating function as ENUM. The enumerating process can be expressed as

$$
\hat{c}_{vis,ENUM} = ENUM(\hat{c}_{vis})
$$

(6)

$$
\hat{c}_{th,ENUM} = ENUM(\hat{c}_{th})
$$

(7)

In the real world, the camera perceives different segments of $\hat{c}_{th,ENUM}$, depending on the view angle. To simulate this effect, we run a random sliding window over $\hat{c}_{th,ENUM}$ with a random location and width. The segment which the sliding window designate is cropped. We define this process as CROP and the cropped segment can be obtained as follows.

$$
\hat{c}_{vis,CROP} = CROP(\hat{c}_{vis,ENUM})
$$

(8)

$$
\hat{c}_{th,CROP} = CROP(\hat{c}_{th,ENUM})
$$

(9)

To consider the deformation of cloth on the non-rigid human body, Thin Plate Spline (TPS) is utilized to approximate this real-world effect.

$$
\hat{c}_{vis,T} = TPS(\hat{c}_{vis,CROP})
$$

(10)

$$
\hat{c}_{th,T} = TPS(\hat{c}_{th,CROP})
$$

(11)

Pixel values of typical digital images do not directly correspond to their captured pixel values in the physical world due to real-world distortions. Expectation over Transformation (EOT) which includes random scaling, illumination adjustment, rotating, and random noising is applied.

$$
\hat{c}_{vis,EOT} = EOT(\hat{c}_{vis,T}, pos, \Theta, k)
$$

(12)

$$
\hat{c}_{th,EOT} = EOT(\hat{c}_{th,T}, pos, \Theta, k)
$$

(13)

For every EOT, same position, rotation angle and scale are applied to $\hat{c}_{vis,T}$ and $\hat{c}_{th,T}$.

Optimization Procedure

In this paper, we attack the state-of-the-art multispectral detector Kim, Park, and Ro (2021) proposed, which is based on Faster R-CNN with a Vgg-16 backbone architecture. This is a typical architecture in which most multispectral detectors
are based on (Liu et al. 2016; Zhang et al. 2019; Li et al. 2018, 2019; Konig et al. 2017). It consists of a two-stage framework: RPN(Region Proposal Network) and a head network. Following existing physical attack methods on two-stage detectors, our designed loss focuses on attacking the RPN and the head network. \( \{ \hat{f}_{BB}^{th}, f_{BB}^{vis} \} \in \mathbb{R}^{H \times W \times C} \) denote the features extracted from each visible and thermal backbone network, and \( \{ S_{vis}^{RPN}, S_{th}^{RPN} \} \in \mathbb{R}^k \) denote the objectness scores of the corresponding anchors (\( k \) total anchors), \( l_{RPN}^{th} \in [0, 1]^k \) denote the ground truth RPN label which indicates presence of an object at the corresponding anchor and \( M^{vis} \) and \( M^{th} \) is the cardinality of nonzero vectors of \( l_{RPN}^{vis} \) and \( l_{RPN}^{th} \). Masking parameters with nonzero vectors of \( l_{RPN}^{th} \) significantly boosts the optimization process. See the Supplementary for details. Our objective of \( \mathcal{L}_{RPN}^{vis} \) is to minimize the objectiveness scores of our target class objects. \( \sigma \) indicates the softmax function.

\[
\mathcal{L}_{RPN}^{vis} = \frac{1}{M^{vis}} \sum_{i=1}^{M^{vis}} \sigma(S_{vis}^{RPN}(f_{BB}^{vis})) \odot l_{RPN}^{vis} \tag{14}
\]

\[
\mathcal{L}_{RPN}^{th} = \frac{1}{M^{th}} \sum_{i=1}^{M^{th}} \sigma(S_{th}^{RPN}(f_{BB}^{th})) \odot l_{RPN}^{th} \tag{15}
\]

\[
\mathcal{L}_{RPN} = \mathcal{L}_{RPN}^{vis} + \mathcal{L}_{RPN}^{th} \tag{16}
\]

In addition to the RPN loss, we add a loss to minimize the classification scores produced by the head network. The visible ROI features and thermal ROI features are concatenated and \( 1 \times 1 \) conv is applied to produce the fused feature \( f_{fuse} \) which is fed to the head network \( H_{cls} \) for final prediction. The head network outputs bounding boxes of objects and corresponding class scores. Likewise, the usage of \( l_{Head} \) significantly boosts the optimization process. See the Supplementary for details.

\[
\mathcal{L}_{Head} = \frac{1}{M} \sum_{i=1}^{M} \sigma(H_{cls}(f_{fuse})) \odot l_{Head} \tag{17}
\]

Finally, \( \mathcal{L}_{patch} \) is designed to ensure the generated pattern is physically realizable. We apply the total variation (TV) Loss, non-printability score (NPS) while generating \( \hat{p}_{vis} \) and \( \hat{p}_{th} \) (Sharif et al. 2016). The TV loss is to make adjacent pixels have similar values to obtain smoother patterns. NPS is adopted such that pixel values of \( \hat{p}_{vis} \) and \( \hat{p}_{th} \) can be expressed by a laser printer and intensity levels by Low-e films. The total loss function to generate \( \hat{p}_{vis} \) and \( \hat{p}_{th} \) can be expressed as the following:

\[
\mathcal{D}_{MS} = \mathcal{L}_{Head} + \lambda_1 \mathcal{L}_{RPN} + \lambda_2 \mathcal{L}_{patch} \tag{18}
\]

We adopt the iterative attack method, similar with the PGD method where \( \alpha \) is the step size, \( i \) is the iteration number, \( \prod \) is the projection function that projects \( \hat{p}_{vis} \) and \( \hat{p}_{th} \) to a feasible set \( P = \{ P : \| P \|_\infty \leq \epsilon \text{ and } A(x, I, P) \in [0, 1]^{H \times W \times 3} \} \)

\[
\hat{p}_{vis}^{i+1} = \prod {p}_{vis}^{i} + \alpha \text{sign}(\nabla p_{vis} \mathcal{D}_{MS}(x_{vis}, x_{th})) \tag{19}
\]

\[
\hat{p}_{th}^{i+1} = \prod {p}_{th}^{i} + \alpha \text{sign}(\nabla p_{th} \mathcal{D}_{MS}(x_{vis}, x_{th})) \tag{20}
\]
driving scenes (1694 image pairs). We compared our attack performance with MAP (Kim, Lee, and Ro 2022), visible-only attack, thermal-only attack denoted as Visible attack and Thermal attack in Figure 5. We evaluated MAP with the same patch size for a fair comparison. Also, we evaluated MAPx2.0 which we scaled up each side length by two times respectively. For the visible-only attack and thermal-only attack, we follow the same optimization procedure except that only one of the *p*_{vis} or *p*_{th} is optimized. For the surveillance dataset we collected, our proposed attack achieves an 82.3% AP drop. Other attacks, Visible attack, Thermal attack, MAP and MAPx2.0 made the AP of the detector drop by 11.2%, 13.5%, 30.6% and 80.9%, respectively. Similarly, our proposed attack achieves 56.8% AP drop under the FLIR-aligned dataset, which consists of driving scenes. Other attacks, Visible attack, Thermal attack, MAP and MAPx2.0 achieve AP drop of 7.4%, 11.6%, 20.0% and 51.2%, respectively. Results show that Visible attack and Thermal attack have less adversarial effect on multispectral detectors. Also, our proposed attack surpasses the attack performance against MAPx2.0 by 3.2%, and 5.6% AP drops on two datasets, with only 1/4 spatial size. Experiment results show that our proposed attack performs superior to other attacks.

### Effect of the Patch Size

We adjust the scale of the original patch to measure the attack performances with different patch sizes.

<table>
<thead>
<tr>
<th>Scale</th>
<th>MAP</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.75</td>
<td>4.5%</td>
<td>54.7%</td>
</tr>
<tr>
<td>1</td>
<td>30.6%</td>
<td>82.3%</td>
</tr>
<tr>
<td>1.25</td>
<td>42.1%</td>
<td>93.9%</td>
</tr>
<tr>
<td>1.5</td>
<td>58.0%</td>
<td>94.2%</td>
</tr>
<tr>
<td>2.0</td>
<td>80.9%</td>
<td>94.3%</td>
</tr>
</tbody>
</table>

Table 2: AP drop across different patch sizes

### Grid Resolution of the Thermal Pattern

Different grid resolutions of the thermal pattern are applied to test the attack strength.

<table>
<thead>
<tr>
<th>Resolution</th>
<th>AP drop</th>
</tr>
</thead>
<tbody>
<tr>
<td>15 × 9</td>
<td>69.0%</td>
</tr>
<tr>
<td>20 × 12</td>
<td>82.3%</td>
</tr>
<tr>
<td>25 × 15</td>
<td>79.2%</td>
</tr>
<tr>
<td>30 × 18</td>
<td>76.7%</td>
</tr>
</tbody>
</table>

Table 3: AP drop with different resolutions of the thermal pattern

### Number of Intensity Levels of the Thermal Pattern

We tested the attack strength across different intensity levels (N).

<table>
<thead>
<tr>
<th>N</th>
<th>AP drop</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>59.2%</td>
</tr>
<tr>
<td>3</td>
<td>82.3%</td>
</tr>
<tr>
<td>5</td>
<td>86.1%</td>
</tr>
<tr>
<td>10</td>
<td>87.0%</td>
</tr>
</tbody>
</table>

Table 4: AP drop with Different Number of Intensity levels (N) composing the thermal pattern
Figure 6: Detection examples of physical attacks. Persons are captured from different angles (top) and distances (bottom). The person wearing the Multispectral Invisible Suit successfully hides from the multispectral detector.

Figure 7: Manufacturing process of the Multispectral Invisible Suit.

Evaluating the Attack in the Physical World

A strong advantage of our proposed physical attack is that it is lightweight, flexible, and self-adhesive such that it can be laminated to any object. By exploiting these properties, we designed a Multispectral Invisible Suit, a wearable clothing that hides persons from multispectral detectors. We requested the tailor to manufacture textured clothing with our visible attack pattern. The tailor periodically enumerated the visible pattern on the clothing so that it covers the full body, including sleeves and the side of the pants, while keeping the size and ratio of the pattern the same as we applied during the digital attack. Then we attached the Low-e films to the clothing according to the optimized thermal pattern. The top and pants consists of total 25 basic patterns of $\hat{p}_{vis}$ and $\hat{p}_{th}$. The manufacturing process of the Multispectral Invisible Suit is briefly illustrated in Figure 7. As shown in Figure 6, the visible attack pattern is well transmitted through the Low-e film, while the designated thermal pattern is observed by the thermal camera. We tested Multispectral Invisible Suit under multiple scenes considering different illumination conditions. We took 30 videos of total of 5400 image frames consisting of different camera view angles and distances from the camera. The person wearing the suit was ordered to rotate counterclockwise at a constant speed at a distance of 6.5 meters from the camera. Also, to evaluate our physical attack from different distances, typical positions between 6 meters and 12 meters are selected and tested. For quantitative evaluation, we use Attack Success Rate, which is defined as the ratio of the number of undetected objects to the total number of objects. Results are shown in Figure 8.

Conclusion

We proposed a new physical attack: Multispectral Invisible Coating. We physically realize a laminated visible-thermal attack using transparent Low-e films. Moreover, we manufactured a Multispectral Invisible Suit that hides persons at multiple views and different distances from the multispectral detector. Extensive experiments show that our physical method effectively attacks multispectral detectors both digitally and physically. Moreover, we collected a new multispectral pedestrian dataset to evaluate our physical attack in surveillance scenes. We expect our physical attack method and the collected dataset we release in public will provide a benchmark for future research on developing robust multispectral detectors.
Acknowledgements

This work was conducted by Center for Applied Research in Artificial Intelligence (CARAI) grant funded by Defense Acquisition Program Administration (DAPA) and Agency for Defense Development (ADD) (UD190031RD).

References


