Information-Theoretic Analysis of Vision-Aided ISAC over a Discrete Memoryless Channel

Xiangliu Tu, Husheng Li, Harpreet S. Dhillon, *Fellow, IEEE*

***Abstract*—We investigate a vision-aided integrated sensing and communications (ISAC) system comprising a transmitter, a receiver and a vision sensor (such as a camera) co-located with the receiver. The vision-aided ISAC system uses the vision sensor to sense the environment and share the vision data with the receiver. The receiver decodes the transmitted message using the received signal and the vision data. Even though this vision data may not completely determine the channel impulse response, some information about the environment, such as whether the transmitter is visible from the receiver, could be potentially useful for decoding. The objective of this paper is to understand the value of such information, termed *channel state knowledge*, using an information-theoretic formalism. We examine three scenarios in which the vision sensor provides different amounts of channel state knowledge to the receiver: perfect, imperfect, and none. Further, we analyze the mutual information for the vision-aided ISAC system using joint and sequential processing approaches and demonstrate that the system with the joint processing of vision data and communication signals has higher mutual information. This analysis provides crucial insights into the performance limits of vision-aided ISAC systems.**

***Index Terms*—Vision-aided communications, discrete memoryless channels, integrated sensing and communications, information theory.**

I. Introduction and Problem Setup

Integrated sensing and communications (ISAC) has emerged as an appealing technology for 6G and beyond wireless networks [1]. A basic ISAC setup consists of a transmitter, a receiver, and a sensor. The transmitter sends an encoded sequence to the receiver through a noisy channel while the receiver tries to decode this sequence from the noisy received signal. Simultaneously, the sensor intercepts a distorted version of the transmitted sequence and seeks to gain insights into channel conditions using this sequence. In many ISAC systems, the notable assumption is that the receiver has access to the perfect channel knowledge, serving as ancillary information for message decoding [1], [2], and the sensor knows the transmitted sequence as side information to estimate the unknown channel conditions. However, attaining the perfect channel knowledge is a formidable challenge in practice.

Since numerous wireless transceivers are co-located with other sensors (such as in vehicular settings), it is highly appealing to use or repurpose existing vision sensors to observe the environment, which presents an alternate way of obtaining some potentially incomplete but still useful channel state information. Therefore, vision sensors have the potential to provide the receiver with valuable channel insights. We term this the *vision-aided ISAC* in this paper. Specifically, the *channel state knowledge* in this paper refers to useful knowledge about physical effects impacting the channel state, such as the placement of the transmitter/receiver and other objects and obstacles. Crucially, this is different from the channel state information (CSI), which is often estimated at the transmitter and/or receiver (depending upon the communication strategy). An example of the channel state knowledge to keep in mind for this paper would be a vision sensor sensing the existence of obstacles around the transmitter and receiver, which will determine whether the channel is line-of-sight (LOS) and non-line-of-sight (NLOS) and then share this information with the receiver. Clearly, the channel state knowledge, as defined in this paper, will impact the CSI but will usually not determine it completely. The vision-aided setup has attracted growing interest and has found new applications in various wireless communication scenarios [3]–[5]. However, there remains a gap in comprehensively understanding the degree of assistance provided by the vision for communication, particularly from the information-theoretic perspective, which is the main inspiration behind this paper.

More specifically, we consider a vision-aided ISAC setup where a transmitter wants to convey messages to the receiver via a state-dependent noisy channel. A vision sensor is incorporated to sense and provide the receiver with channel state knowledge. This scenario forms the basis for our investigation. Particularly, we consider a discrete memoryless channel without feedback. The setup is depicted in Fig. 1. The transmitter

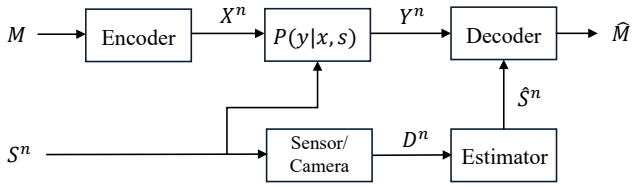


Fig. 1. The illustration of the vision-aided ISAC setup.

encodes the message M into a codeword $X^n = \{X_1, ..., X_n\}$ and transmits it over the channel in n channel uses. The state sequence $S^n = \{S_1, ..., S_n\}$ is independent and identically distributed (i.i.d) according to probability distribution $p(s)$ and S_i is fixed at each transmission but unknown to the sensor and receiver. The vision sensor provides this state knowledge to the receiver through vision data D, aiming to reduce the

X. Tu and H. S. Dhillon are with Wireless@VT, Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, 24061, USA. Email: {xiangliutu, hdhillon}@vt.edu. H. Li is with the School of Aeronautics and Astronautics and the Elmore School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, 47907, USA. Email: husheng@purdue.edu. The support of the US NSF (Grant CNS-2225511) is gratefully acknowledged.

uncertainty about the channel state. With access to the data D, we consider two distinct approaches for processing the communication signal Y and data D at the receiver. One method uses a state estimator to detect and share the estimated state $\hat{S}$ with the receiver. Upon receiving a noisy sequence Y^n and estimated state $\hat{S}^n$, the receiver generates an estimate $\hat{M}$ of the original message M. Alternatively, one can share the vision data with the receiver and jointly process the data D^n with the communication signal Y^n at the decoder. This paper will study the mutual information for both these approaches within this vision-aided ISAC setup.

Like the typical ISAC system, the setup outlined here serves two primary purposes: transmitting the message M and acquiring the channel state knowledge S. However, the key distinction lies in the assumption of many ISACs that perfect channel knowledge is available at the receiver. In this vision-aided ISAC setup, we examine a more realistic scenario where the receiver uses a vision sensor to obtain the channel state knowledge and integrates it into message decoding. Ultimately, the vision-aided ISAC system aims to enhance communications performance by leveraging the vision sensor.

A. Prior Art

A typical ISAC setup uses one signal to convey messages while sensing the target. Therefore, the design of the transmitted signal affects both sensing and communication performance. However, the optimal waveform for communication may not be the same for sensing and vice versa [1], [6]. Therefore, numerous studies have been dedicated to understanding their fundamental trade-offs and assessing the performance limits of the ISAC system, especially from an information-theoretic perspective [7]–[10]. The rate-distortion-cost tuple is a standard metric used to evaluate the performance of an ISAC system. However, the sensing performance evaluated by the rate-distortion tuple varies with different distortion measures. Thus, there has been significant interest in constructing a universal metric for sensing that is not affected by the choice of distortion measure. One proposed metric is the "sensing rate" or "sensing mutual information" [11], also adopted in this paper. The use of mutual information as a metric for sensing can be traced back to the radar waveform design [12]. Recently, more work has incorporated this sensing metric into their analyses, such as [8], where the sensing rate between the state sequence and the channel output is used to measure the extractable state information. The authors in [13] characterize the optimal trade-off between communication and sensing by logarithmic loss distortion and provide one operational meaning of the sensing rate. The study of [14] shows that the sensing mutual information provides a universal lower bound for the distortion metrics of sensing.

Vision-aided communications has gained increasing interest and found new applications in various communication scenarios. For example, the authors of [4] explore millimeter wave (mmWave) communication systems equipped with cameras at the base station. They leverage deep learning tools to directly predict mmWave beams and blockages from the camera RGB images. Similarly, the authors of [15] use a sensor to guide digital beamforming in a multiple-input multiple-output (MIMO) system, forming a dual-sensing setup. Additionally, the authors of [16] design a machine learning framework that uses sensing information to efficiently select optimal reflection beams in a reflecting intelligent surface. The work in [17] jointly processes sequences of vision and wireless data frames to identify the communication user from the other candidate objects. However, most analyses of ISAC setup are constrained by the assumption of the perfect state available at the receiver, which is impractical when analyzing a vision-aided ISAC system. Therefore, we introduce a new evaluation framework using information theory to quantify the benefits of vision analytically and gain insights into the performance limits of vision-aided ISAC systems.

B. Contributions

In this paper, we quantify the degree of assistance the vision sensor provides to communication in a vision-aided ISAC system. We consider three scenarios involving varying amounts of channel state knowledge and analyze the mutual information of these vision-aided ISAC systems. As expected, our analysis shows that more accurate state information provided by the vision sensor leads to better communication performance at the receiver, and the performance improvement is quantified by the mutual information $I(X;\hat{S}|Y)$. The availability of both communication signal and vision data at the receiver presents another important task: finding the optimal way to process these received data. We propose two general processing approaches: one is sequential processing, which utilizes a state estimator to detect the state before decoding, and another is joint processing of vision data and communication signal at the decoder. We demonstrate that joint processing provides more information than the sequential one. The performance difference between the two approaches is quantified by $I(S;(X,Y)|D)$. Most importantly, through the analysis of the joint processing, we also establish that the performance boundary of the vision-aided ISAC system is described by $I(X,S;Y,D)$, which is the mutual information between the pair of the signal-channel (X,S) and pair of observations (Y,D).

II. System Model

We consider a state-dependent discrete memoryless communication channel with input X, output Y, and state S. The receiver observes Y through channel $p(y|x,s)$, a collection of the conditional probability mass functions on $\mathcal{Y}$. The transmitter uses channel n times to send X^n to the receiver. The channel is memoryless in the sense that, without feedback, the joint distribution is decomposable as $p(y^n|x^n,s^n) = \prod_{i=1}^{n} p_{Y|X,S}(y_i|x_i,s_i)$. A vision sensor senses the channel and then provides either perfect or imperfect state knowledge to the decoder. We design three scenarios with different amounts of channel state knowledge to better understand this problem: (1) the perfect case where the vision sensor can always provide the receiver with perfect channel state knowledge S. (2) The imperfect case where the channel state

knowledge provided by the vision sensor is imperfect. (3) The blind case where no state knowledge is available at the receiver. We are interested in comparing the channel capacity of these three cases to illustrate the degree of assistance provided by the vision sensor under these assumptions.

1) Perfect Case: In this case, we assume that the vision sensor shares the perfect states S with the receiver at each transmission. The setup is depicted in Fig. 2. Let $p(y|x) = \sum_s p(y|x,s)p(s)$ be the discrete memoryless channel obtained by averaging the $p(y|x,s)$ over all states. Therefore, the channel capacity for the perfect case is

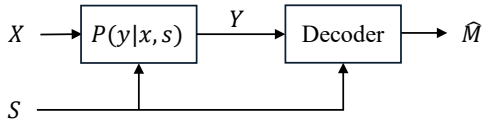


Fig. 2. The illustration of the perfect case where the vision sensor provides perfect state S to the receiver.

$$C = \max_{p(x)} I(X;Y|S), \tag{1}$$

where $p(x)$ is the distribution of input X. Please refer to [18] for its proof.

2) Imperfect Case: This is a highly realistic case where the vision information obtained by the sensor might be noisy because of limited resolution and/or estimation errors. Denoting the imperfect state as $\hat{S}$, the setup is illustrated in Fig. 3. Then,

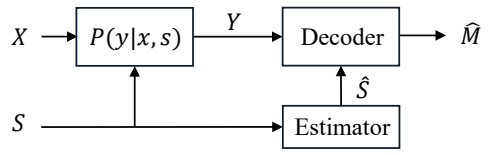


Fig. 3. The illustration of imperfect state knowledge $\hat{S}$ is available to the receiver.

the channel capacity is given as [18]

$$C = \max_{p(x)} I(X;Y|\hat{S}). \tag{2}$$

3) Blind Case: Without access to the state knowledge, the receiver only obtains the output of the communication channel. This scenario is illustrated in Fig. 4. The channel capacity is

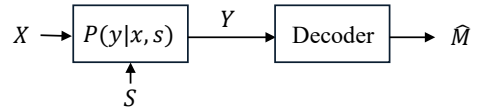


Fig. 4. In the blind case, no state information is available to the receiver.

expressed as [19]

$$C = \max_{p(x)} I(X;Y). \tag{3}$$

Considering the memoryless nature of the channel, the input X and state S are independent. Therefore, we conclude that $I(X;Y|S) \geq I(X;Y)$ and $I(X;Y|\hat{S}) \geq I(X;Y)$. Moreover, due to $I(X;S|Y) \geq I(X;\hat{S}|Y)$, it follows that $I(X;Y|S) \geq I(X;Y|\hat{S})$. Accordingly, we can establish an inequality among three mutual information values as follows

$$I(X;Y|S) \geq I(X;Y|\hat{S}) \geq I(X;Y). \tag{4}$$

Furthermore, we can analytically characterize the performance improvement by the equality:

$$I(X;Y|S) + I(X;S) = I(X;Y) + I(X;S|Y). \tag{5}$$

The independence between X and S indicates $I(X;S) = 0$. Therefore, the difference between the mutual information of perfect and blind cases is given by

$$I(X;Y|S) - I(X,Y) = I(X;S|Y). \tag{6}$$

In the next section, we study a binary symmetric setup to derive explicit expressions for the mutual information and illustrate the effectiveness of this information-theoretic evaluation framework for this vision-aided ISAC system.

A. Binary Symmetric Setup

The binary symmetric setting consists of the transmission indicator X as input with $X \sim \text{Bernoulli}(a)$, where a is the transmission probability. The channel state $S \sim \text{Bernoulli}(b)$, where b denotes the probability of $S = 1$. The independent additive noise is $Z \sim \text{Bernoulli}(p)$, where p is the probability of bit flip in a given trial. The system model is illustrated in Fig. 5. Consequently, the output of this channel is given by

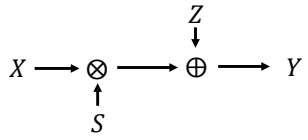


Fig. 5. The illustration of binary symmetric state-dependent channel setting.

$$Y = SX \oplus Z, \tag{7}$$

where summation $\oplus$ denotes mod 2 addition and the output $Y \in \mathcal{Y} = \{0,1\}$. The transition probability of this state-dependent channel is given by

$$P_{Y|X,S}(y|x,s=0) = \begin{cases} 1-p, & y=0 \\ p, & y=1, \end{cases} \tag{8}$$

$$P_{Y|X,S}(y|x,s=1) = \begin{cases} 1-p, & y=x \\ p, & y \neq x. \end{cases} \tag{9}$$

Next, we will analyze the mutual information of perfect, imperfect and blind cases, respectively.

1) Perfect case: Assuming the perfect state S is available at the receiver, from equation (1), the mutual information is shown in the next Proposition.

Proposition 1. *The mutual information for the perfect case is given as*

$$I(X;Y|S) = b\left[(1-a)D(p||Q) + aD(p||1-Q)\right], \tag{10}$$

where $Q = a + p - 2ap$ and $D(\cdot||\cdot)$ is the binary relative entropy defined as $D(p||q) = p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q}$.

Proof: With the definition of the mutual information, we have

$$I(X;Y|S) = \sum_{(x,y,s)} P(x,y,s) \log \frac{p(x,y|s)}{P(x|s)P(y|s)} \tag{11}$$

$$\overset{(a)}{=} \sum_{(x,y,s)} P(x,y,s) \log \frac{P(y|x,s)}{P(y|s)} \tag{12}$$

$$\begin{aligned}\overset{(b)}{=}&(1-p)(1-a)b \log \frac{1-p}{1-p-a+2ap} \\ &+ pab \log \frac{p}{1-p-a+2ap} \\ &+ p(1-a)b \log \frac{p}{a(1-p)+(1-a)p} \\ &+ (1-p)ab \log \frac{1-p}{a(1-p)+(1-a)p} \\ =&(1-a)bD(p||a+p-2ap) \\ &+ abD(p||1-a-p+2ap),\end{aligned} \tag{13}$$

where the step (a) follows from the independence between X and S. Step (b) follows from algebraic manipulations of $P(x=i,y=j,s=k) \log \frac{P(y=j|x=i,s=k)}{P(y=j|s=k)}$ using (8) and (9), for $i,j,k \in \{0,1\}$. For example, $P(x=0,y=0,s=0)=0$ and $P(x=0,y=0,s=1)=(1-p)(1-a)b \log \frac{1-p}{1-p-a+2ap}$. Then, substituting them into (12), we get (13). ∎

Proposition 1 shows that $I(X;Y|S)$ is a linear function in b, multiplied by a convex summation of two binary relative entropy terms. It is easy to show that $I(X;Y|S)$ is convex in p for a fixed a and is concave in a for a fixed p. Furthermore, the channel capacity for the perfect case is achieved when $b=1$, the transmission probability $a=\frac{1}{2}$ and the probability of bit flip $p=0$ or 1.

2) Imperfect case: In the imperfect case, we model S and $\hat{S}$ as the input and output of a binary symmetric channel with the probability of state error $P_e = q$ as illustrated in Fig. 6. Assuming that $\hat{S}$ is available at the receiver, the mutual

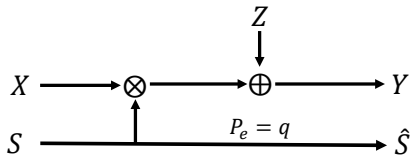


Fig. 6. The illustration of binary symmetric state-dependent channel setting with imperfect state $\hat{S}$ available at the receiver.

information for this case is given in the next Proposition.

Proposition 2. *The mutual information for the imperfect case is given as*

$$\begin{aligned}I(X;Y|\hat{S}) =&(1-a)H(p) - P(\hat{s}=0)H(A) \\ &- P(\hat{s}=1)H(B) \\ &+ P(\hat{s}=1)P(x=1)H(C) \\ &+ P(\hat{s}=1)P(x=0)H(D),\end{aligned} \tag{14}$$

where $A = P(y=0|\hat{s}=0), B = P(y=0|\hat{s}=1), C = P(y=0|x=1,\hat{s}=1), D = P(y=0|x=1,\hat{s}=0)$ *and* $P(\hat{s}=0) = 1-b-q+2bq$.

Proof: From the definition of mutual information, we have

$$I(X;Y|\hat{S}) = \sum_{(x,y,\hat{s})} P(x,y,\hat{s}) \log \frac{P(x,y,\hat{s})}{P(x)P(y,\hat{s})}. \tag{15}$$

Following similar algebraic manipulations as in the proof of Proposition 1, we calculate $P(x,y,\hat{s}) \log \frac{P(x=i,y=j,\hat{s}=k)}{P(x=i)P(y=j,\hat{s}=k)}$, for all $i,j,k \in \{0,1\}$. For example, given $x=0$, $y=0$ and $s=0$, we can calculate

$$\begin{aligned}P(y=0,x=0,\hat{s}=0) =&(1-p)(1-a)(1-b-q+2bq), \\ P(y=0,\hat{s}=0) =&(1-p)(1-a)(1-b-q+2bq) \\ &+ (1-p)a(1-q)(1-b) + paqb,\end{aligned}$$

and $P(y=0|x=0,\hat{s}=0) = 1-p$. Then, substituting them into (15), we obtain (14). ∎

The imperfect case is more intricate than the perfect case. Thus, we provide specific instances to understand this case better. For example, the mutual information of the imperfect case is equal to the perfect case $I_{q\in\{1,0\}}(X;Y|\hat{S}) = I(X,Y|S)$ when estimated state $\hat{S}$ is the same as the perfect state S or its inverse, for $q=1$ or 0. Also, the imperfect case has the same performance as the blind case $I_{q=0.5}(X;Y|\hat{S}) = I(X;Y)$ when $P(\hat{S}=S)=\frac{1}{2}$ which indicates no useful state knowledge available at the receiver. In addition, $I(X;Y|\hat{S})$ exhibits the identical convexity property as $I(X;Y|S)$ for the fixed state parameter b and probability of state error P_e.

3) Blind Case: Now, we consider the scenario where no information about the state is available. Mutual information for the blind case is given in the Proposition 3.

Proposition 3. *The mutual information for the blind case is given as*

$$\begin{aligned}I(X;Y) =&(1-a)D(p||P(y=1)) \\ &+ aD(p+b-2pb||P(y=1)),\end{aligned} \tag{16}$$

where $P(y=1) = p+ab-2abp$.

Proof: For the blind case, we calculate $P(x,y) \log \frac{P(x=i,y=j)}{P(x=i)P(y=j}$, for all $i,j \in \{0,1\}$ in the same way of other two cases. Then, the mutual information is given as

$$\begin{aligned}I(X;Y) =& \sum_{(x,y)} P(x,y) \log \frac{p(x,y)}{P(x)P(y)} \\ =&(1-p)(1-a) \log \frac{1-p}{P(y=0)} \\ &+ (1-p-b+2pb)) \log \frac{1-p-b+2pb}{P(y=0)} \\ &+ p(1-a)b \log \frac{p}{P(y=1)} \\ &+ (p+b-2pb) \log \frac{p+b-2pb}{P(y=1)} \\ =&(1-a)D(p||P(y=1)) \\ &+ aD(p+b-2pb||P(y=1)),\end{aligned} \tag{17}$$

which completes the proof. ∎

Specifically, the mutual information of the blind case is the same as that of the perfect case when the state parameter $b=1$

and we have $I_{b=1}(X;Y|S) = I_{b=1}(X;Y)$. This observation can be explained by the fact that when the channel state is fixed, the state knowledge becomes inconsequential to the receiver. However, incorporating state knowledge is crucial for practical settings where channel states continuously vary over time. Besides, $I(X;Y)$ shows the same convexity property as $I(X;Y|S)$ for a fixed b.

III. Sequential and Joint Processing

In the vision-aided ISAC system, when the transmitter conveys messages to the receiver, the vision sensor simultaneously senses and shares the vision data about the state with the receiver. Upon receiving the vision data D and communication signal Y, there are two possibilities for processing them. One is *sequential processing*, which employs a state estimator to detect and provide the estimated state $\hat{S}$ to the receiver as depicted in Fig. 7(a). The receiver then decodes the message by taking $\hat{S}$ as the real state. However, the sequential processing of Y and D may result in a loss of information. Hence, we also direct our attention to the *joint processing* as illustrated in Fig. 7(b), where the vision data is delivered to the decoder directly without any prior state estimation. This section aims to study mutual information for both these approaches to gain a deeper understanding of their differences. Vision data is used to reduce uncertainty about the channel state. Therefore, utilizing mutual information $I(S;D)$ to quantify the information conveyed by D about the state S is reasonable.

1) Sequential Processing: In Fig. 7(a), the dotted box frames the pieces for sensing the channel state, which can be viewed as a sensing channel to provide state information. The uncertainty reduction about the state is quantified by the sensing rate Δ, which is defined by the mutual information

$$\Delta \leq I(S;D). \tag{18}$$

The box with a solid line around it frames the pieces for communication, and the communication rate R is defined as

$$R \leq I(X;Y|\hat{S}). \tag{19}$$

The quantities R and Δ share the same units when utilizing the same logarithmic base, such as the commonly used base 2. The communication rate R quantifies the uncertainty reduction about X given the signal Y. Therefore, a vision-aided ISAC system under sequential processing can be viewed as a combination of two parallel channels. Accordingly, it is reasonable to define the mutual information of the vision-aided ISAC system using sequential processing as the summation of the mutual information of the two channels

$$\begin{aligned} R+\Delta \leq & I(S;D) + I(X;Y|\hat{S}) \\ \overset{(a)}{\leq} & I(S;D) + I(X;Y|D), \end{aligned} \tag{20}$$

where step (a) follows by the Markov chain $S \to D \to \hat{S}$ and the data processing inequality.

2) Joint Processing: The sequential approach processes Y and D separately, inevitably resulting in a potential loss of information. Therefore, we now investigate what happens if we process them jointly. The mutual information of the vision-aided ISAC system with the joint processing is given in Theorem 1.

Theorem 1. *The mutual information for vision-aided ISAC system with a state-dependent memoryless channel $(\mathcal{X} \times \mathcal{S}, p(y|x,s), \mathcal{Y},)$ using the joint processing satisfies*

$$R + \Delta \leq I(X,S;Y,D), \tag{21}$$

where $I(X,S;Y,D)$ is the mutual information between the signal-channel pair (X,S) and observation pair (Y,D).

Proof: The proof of Theorem 1 consists of two parts. The first is the proof of achievability. We need to demonstrate that any pair of (R,Δ) satisfying (21) for some $p(x)$ and $p(s)$ is achievable. For fixed $p(x)$, [18] shows that the transmitter can send $I(X;Y|D)$ bits reliably across the channel, and the sensor can reduce $I(S;D)$ bits uncertainty about state via sensing channel for fixed $p(s)$. With high probability, the receiver can decode the codeword X from the received signal Y. Therefore, an additional state uncertainty reduction is from $H(S|D)$ to $H(S|X,Y,D)$. Accordingly, the uncertainty reduction of channel state S is given as

$$\begin{aligned} \Delta &= I(S;D) + H(S|D) - H(S|X,Y,D) \\ &= I(S;D) + I(S;(X,Y)|D). \end{aligned} \tag{22}$$

As a result, the mutual information for the joint approach satisfies

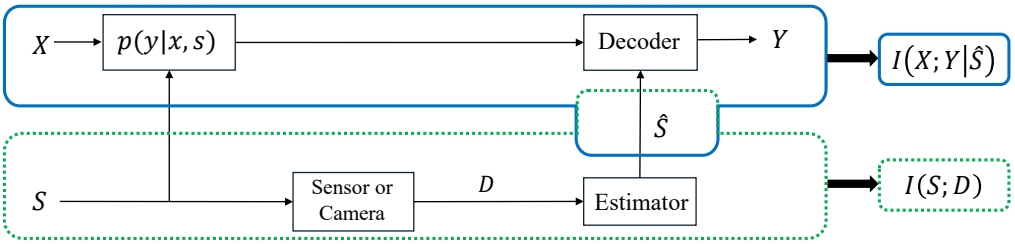
$$\begin{aligned} R+\Delta &\leq I(X;Y|D) + I(S;D) + I(S;(X,Y)|D) \\ &= I(X,S;Y,D), \end{aligned} \tag{23}$$

for any fixed $p(x)$ and $p(s)$ is achievalble.

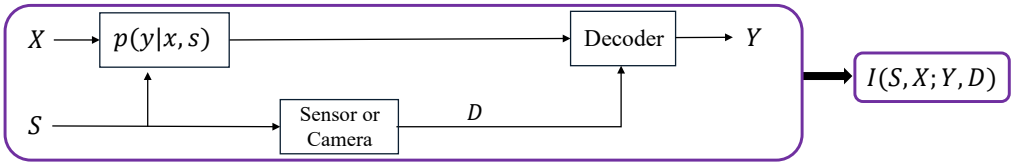
For the proof of converse, we need to show that for every sequence of $(2^{nR},n)$ codes with $\lim_{n\to\infty} P_e^{(n)} = 0$, it must have $R \leq \max_{p(x)} I(X;Y|D)$. Followed by the standard steps in [18, Ch.3], we bound the communication rate by

$$\begin{aligned} nR =& H(M) \\ =& I(M;Y^n,D^n) + H(M|Y^n,D^n) \\ \overset{(a)}{\leq}& I(M;Y^n,D^n) + n\epsilon_n \\ \overset{(b)}{=}& \sum_{i=1}^{n} I(M;Y_i,D_i|Y^{i-1},D^{i-1}) + n\epsilon_n \\ \leq& \sum_{i=1}^{n} I(M,Y^{i-1},D^{i-1};Y_i,D_i) + n\epsilon_n \\ \overset{(c)}{\leq}& \sum_{i=1}^{n} I(X_i,Y^{i-1},D^{i-1};Y_i,D_i) + n\epsilon_n \\ \overset{(d)}{=}& \sum_{i=1}^{n} I(X_i;Y_i,D_i) + n\epsilon_n, \end{aligned} \tag{24}$$

where ϵ_n tends to zero as $n \to 0$. Step (a) follows the Fano's inequality $H(M|Y^n) \leq n\epsilon_n$ [18, Ch.3.1], step (b) follows by the chain rule of mutual information, step (c) is



(a) Illustrating vision-aided ISAC system system with sequential processing.



(b) The illustration of vision-aided ISAC system with joint processing.

Fig. 7. (a) The sequential processing involves using a state estimator, which takes the vision data as input and provides the estimated state $\hat{S}$ to the receiver. (b) In joint processing, the vision sensor shares the vision data with the receiver directly without prior state estimation.

an outcome of data processing inequality and $X^n = g(W)$, where $g(\cdot)$ is an encoder function, and step (d) follows from the memoryless assumption of the channel. By leveraging the concavity of the mutual information in the input distribution, Jensen's inequality and the independence between data D and X, we have $R \leq I(X;Y|D) + \epsilon_n$.

Similarly, we bound the rates summation $R + \Delta$ by

$$\begin{aligned} n(R+\Delta) =& H(M) + H(S) \\ \leq& I(M;Y^n) + I(S^n;D^n) + n\epsilon_n' \\ \overset{(a)}{\leq}& I(M;Y^n|D^n) + I(S^n;D^n|X^n) + n\epsilon_n' \\ =& I(M;Y^n|D^n) + I(S^n;D^n) \\ &+ I(S^n;X^n|D^n) + n\epsilon_n' \\ \overset{(b)}{\leq}& I(M;Y^n|D^n) + I(S^n;D^n) \\ &+ I(S^n;X^n,Y^n|D^n) + n\epsilon_n' \\ =& I(X^n;Y^n|D^n) + I(S^n;D^n) \\ &+ I(S^n;X^n,Y^n|D^n) + n\epsilon_n' \\ \overset{(c)}{=}& \sum_{i=1}^{n} I(S_i,X_i;Y_i,D_i) + n\epsilon_n', \end{aligned} \tag{25}$$

where ϵ_n' tends to zero as $n \to 0$. Step (a) follows from the fact that M and X^n are independent of S^n, and conditioning on D and X increases mutual information. Step (b) follows from $I(S^n;X^n,Y^n|D^n) \geq I(S^n;X^n|D^n)$ and step (c) follows from the memoryless channel assumption. Finally, we have $R + \Delta \leq I(S,X;Y,D) + \epsilon_n'$, which completes the proof. ■

To compare the performance of the sequential and joint approaches, we define the maximum mutual information for vision-aided ISAC using sequential processing as $I_{\text{Sequential}} = I(S;D) + I(X;Y|D)$ and for the joint processing as $I_{\text{Joint}} = I(X,S;Y,D)$. From (23), we have $I_{\text{Joint}} = I_{\text{Sequential}} + I(S;(X,Y)|D)$, where $I(S;(X,Y)|D) \geq 0$. Consequently, we conclude that

$$I_{\text{Joint}} \geq I_{\text{Sequential}}. \tag{26}$$

Equality is achieved at $D = S$ when the vision sensor provides the perfect channel state S. The intuitive explanation is that when S is available, the reduction in uncertainty only stems from the communication channel, which is the same for both of these approaches.

IV. Numerical Illustrations

This section provides visual representations of our main results and conclusions about the vision-aided ISAC system with a binary symmetric setting. First, we show mutual information as a function of the probability of transmission a for three cases. The Fig. 8 shows the mutual information values for three cases are the same, given the state parameter $b = 1$. The mutual information decreases as the probability of bit flip p increases. For the concavity, mutual information is the summation of two log functions of a, given fixed b, p and q. Therefore, it is concave. Then, we compare the mutual information of three cases with different state parameters b

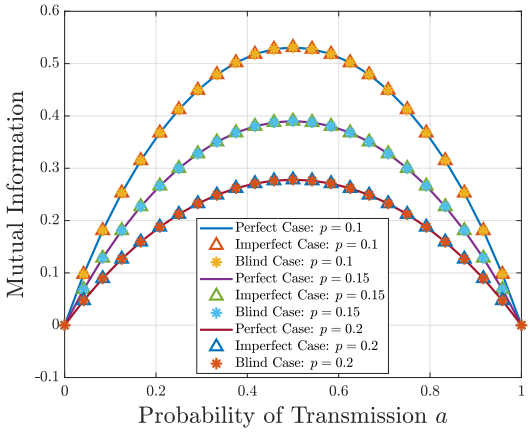


Fig. 8. The mutual information of perfect, imperfect and blind case for $b = 1$, $p = 0.1, 0.15$ and 0.2, and $q = 0.9$.

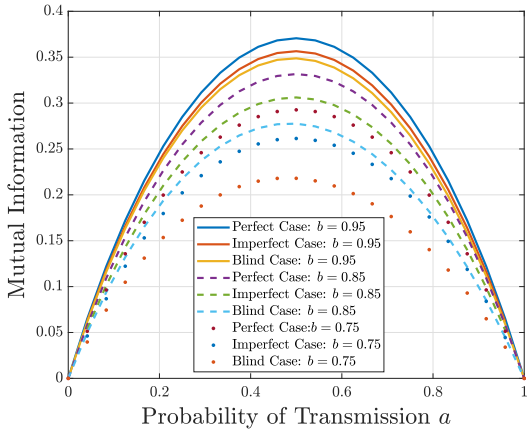


Fig. 9. The mutual information of perfect, imperfect and blind case for $b = 0.95, 0.85$ and 0.75, $p = 0.1$ and $q = 0.9$.

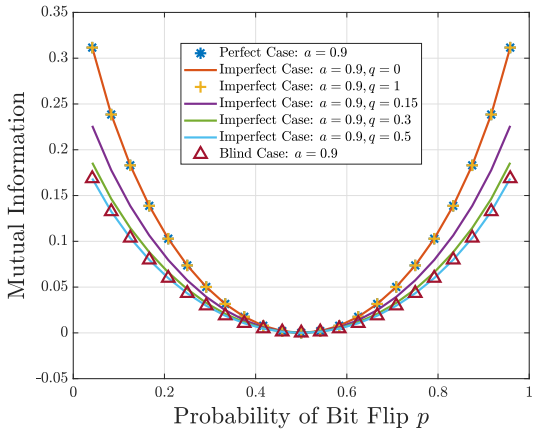


Fig. 10. The mutual information of perfect, imperfect and blind case for $a = 0.9$ and $q = 0, 0.15, 0.3, 0.5$ and 1 .

in Fig. 9, given the probability of bit flip $p = 0.1$ and the probability of state error $q = 0.9$. The results and observations are consistent with our conclusion in (4). Additionally, we show the mutual information as a function of the probability of bit flip p in Fig. 10, given different probabilities of state error q. The mutual information of the imperfect case is the same as that of the perfect case when $q = 0$ or 1, while $q = \frac{1}{2}$, the imperfect case has the same mutual information as the blind case since the state uncertainty is maximized under this scenario. The mutual information increases as the probability of state error decreases.

V. Conclusion

A vision-aided ISAC system utilizes vision sensors to provide valuable channel state knowledge to the communications receiver. In this paper, we have investigated this vision-aided ISAC setup, shedding light on the degree of assistance that vision provides to communication from an information theory perspective. We have considered three scenarios with different amounts of channel state knowledge available at the receiver and showed that more accurate state knowledge received from the vision sensor increases the mutual information at the receiver. The performance improvement is analytically quantified by mutual information $I(X; S|Y)$. We have also proposed sequential and joint approaches for the receiver to process communication and vision data and demonstrated that the joint approach is better for a vision-aided ISAC system. In this analysis, we have implemented the sensing rate to quantify the state uncertainty reduction provided by a vision sensor and show that the performance limit of the vision-aided ISAC system is characterized by $I(X, S; Y, D)$. This work has numerous extensions possible. As the immediate next step, we are relaxing the i.i.d. channel assumption for a possible journal extension of this work.

References

[1] F. Liu, Y. Cui, C. Masouros, J. Xu, T. X. Han, Y. C. Eldar, and S. Buzzi, "Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond," *IEEE Journal on Sel. Areas in Commun.*, vol. 40, no. 6, pp. 1728–1767, 2022.

[2] A. Liu, M. Li, M. Kobayashi, and G. Caire, "Fundamental limits for ISAC: Information and communication theoretic perspective," in *Integrated Sensing and Commun.* Springer, 2023, pp. 23–52.

[3] G. Charan, M. Alrabeiah, and A. Alkhateeb, "Vision-aided dynamic blockage prediction for 6G wireless communication networks," in *Proc., IEEE Intl. Conf. on Commun. (ICC)*. IEEE, 2021, pp. 1–6.

[4] M. Alrabeiah, A. Hredzak, and A. Alkhateeb, "Millimeter wave base stations with cameras: Vision-aided beam and blockage prediction," in *Proc., IEEE Veh. Technology Conf. (VTC)*. IEEE, May 2020, pp. 1–5.

[5] J. Bao, T. Shu, and H. Li, "Handover prediction based on geometry method in mmWave communications - A sensing approach," in *Proc., IEEE Intl. Conf. on Commun. (ICC)*, 2018, pp. 1–6.

[6] Y. Xiong, F. Liu, K. Wan, W. Yuan, Y. Cui, and G. Caire, "From torch to projector: Fundamental tradeoff of integrated sensing and communications," *IEEE BITS the Info. Theory Magazine*, Mar. 2024.

[7] M. Ahmadipour, M. Kobayashi, M. Wigger, and G. Caire, "An information-theoretic approach to joint sensing and communication," *IEEE Trans. on Info. Theory*, 2022.

[8] Y.-H. Kim, A. Sutivong, and T. M. Cover, "State amplification," *IEEE Trans. on Info. Theory*, vol. 54, no. 5, pp. 1850–1859, 2008.

[9] N. R. Olson, J. G. Andrews, and R. W. Heath, "Coverage and rate of joint communication and parameter estimation in wireless networks," *IEEE Trans. on Info. Theory*, vol. 70, pp. 206–243, 2024.

[10] H. Joudeh and F. M. Willems, "Joint communication and binary state detection," *IEEE Journal on Selected Areas in Information Theory*, vol. 3, pp. 113–124, 2022.

[11] C. Ouyang, Y. Liu, H. Yang, and N. Al-Dhahir, "Integrated sensing and communications: A mutual information-based framework," *IEEE Commun. Magazine*, vol. 61, no. 5, pp. 26–32, 2023.

[12] M. R. Bell, "Information theory and radar waveform design," *IEEE Trans. on Info. Theory*, vol. 39, no. 5, pp. 1578–1597, 1993.

[13] H. Joudeh and G. Caire, "Joint communication and state sensing under logarithmic loss," in *IEEE Intl. Symposium on Joint Commun. & Sensing (JC&S)*. Institute of Electrical and Electronics Engineers, Mar. 2024.

[14] F. Liu, Y. Xiong, K. Wan, T. X. Han, and G. Caire, "Deterministic-random tradeoff of integrated sensing and communications in Gaussian channels: A rate-distortion perspective," in *Proc., IEEE Intl. Symposium on Information Theory*. IEEE, May 2023, pp. 2326–2331.

[15] S. Jiang, A. Alkhateeb, D. W. Bliss, and Y. Rong, "Vision guided MIMO radar beamforming for enhanced vital signs detection in crowds," *IEEE Trans. on Aerospace and Electronic Systems*, 2024.

[16] S. Jiang, A. Hindy, and A. Alkhateeb, "Camera aided reconfigurable intelligent surfaces: Computer vision based fast beam selection," in *Proc., IEEE Intl. Conf. on Commun. (ICC)*. IEEE, Oct. 2023, pp. 2921–2926.

[17] G. Charan and A. Alkhateeb, "User identification: A key enabler for multi-user vision-aided communications," *IEEE Open Journal of the Commun. Society*, 2023.

[18] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge university press, 2011.

[19] T. M. Cover, *Elements of information theory*. John Wiley & Sons, 1999.