Lossy compression of active sources

Hari Palaiyanur, Cheng Chang and Anant Sahai
Department of Electrical Engineering and Computer Sciences
University of California at Berkeley
Berkeley, CA 94720
{hpalaiya, cchang, sahai}@eecs.berkeley.edu

***Abstract*— In computer vision, an active vision source is a sensor that explores its environment in an active way, deciding to investigate parts of the environment in greater depth based on what it currently sees. We study the problem of determining the rate required to compress the output of an active vision source to within a desired fidelity. In order to make the problem analytically tractable, we assume that the environment is memoryless and gain insights into the distinction between compression of passive and active sources. We show that modelling of the sources is crucial by considering two extreme cases: adversarially active sources and helpful active sources. The theory of arbitrarily varying sources is useful for these purposes and we expand on it by allowing the party controlling the variation in the source to have partial or noisy observations of the environment. We give several examples showing that there is a large difference in the rate required to compress active sources that are adversarially modelled and active sources that are jointly optimized with the coding system. The results suggest that when active sources are part of a networked system where rate comes at a premium, large savings can be reaped by jointly optimizing the coding system with the computer vision system.**

I. Introduction

Active vision/sensing/perception [1] is an approach to computer vision, the main principle of which is that sensors should choose to explore their environment actively *based on what they currently sense or have previously sensed*. As Bajcsy states it in [1], "We do not just see, we look." The contrast between active and passive vision is easily understood through examples. Consider a fixed security camera. This camera is passive and does not have the ability to change position or optically zoom in on different areas, it merely records. On the other hand, consider a camera that has the ability to rotate in several dimensions and can optically zoom. This active camera can be operated autonomously using adaptive rules or by a human to gain greater resolution of particular objects in the environment, e.g. trespassers in a building or endangered animals in the wild.

Active vision sources may also have noncausal knowledge of the environment. For example, compare the recordings of a cameraman filming a wild animal and a cameraman filming a movie. When filming a wild animal, fairly little is predictable and the focus of the camera is dependent almost entirely on what has happened and what is currently happening. When filming a movie, however, the cameraman has access to a script and knows ahead of time which parts of the environment are to be filmed. Noncausality can also be thought of as a way of abstracting the ability of the sensor to predict the future based on previous observations. Even amongst causal active sources, there is a distinction between causal and strictly causal sources. This distinction is mostly related to the time-scales on which the environment changes and the sensor can actively sense and move. If the environment changes slowly enough that knowledge of the immediate past allows for prediction of the present (and actuation by the sensor within the necessary time), it is as if the sensor knows the present. If this is not true, however, there is intuitively a large difference between the sensor knowing the past (strictly causal) and the sensor knowing the present (causal).

This work was supported by a NSF Graduate Research Fellowship.

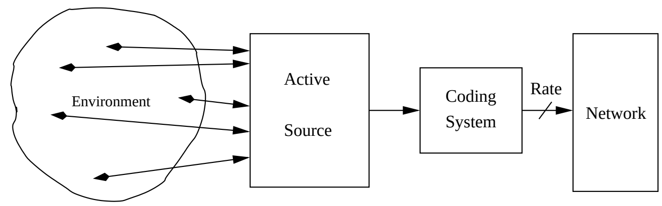

Fig. 1. The problem of interest: lossy compression of an active source that dynamically samples its environment.

In this paper, we are interested in the fixed-rate lossy compression of an active vision source. Such a problem arises when active sources are embedded in a networked system and their rate requirements must be budgeted. In order to fully understand this problem, there are many individual questions that need to be answered. For example, we would need to know the relevant distortion criteria for encoding of video. Also, how should we model the plenoptic function [2] the active source samples? For analytical tractability, we assume these issues are dealt with and consider an admittedly oversimplified model using known distortion measures, finite alphabets[1], stationary, memoryless environments and active agents that observe the environment through a discrete memoryless channel (DMC). Our basic question is the following: How does the rate-distortion function change when a party can provide the coding system with input based on noisy or partial knowledge of the environment. Furthermore, we ask how important modelling of this agent is in determining the rate-distortion function.

In general, the active vision source has an objective that

[1]The finite alphabet restriction is reasonable because most sensors have a digital interface with the analog world. While the alphabets would certainly be unfathomably large, they are still finite.

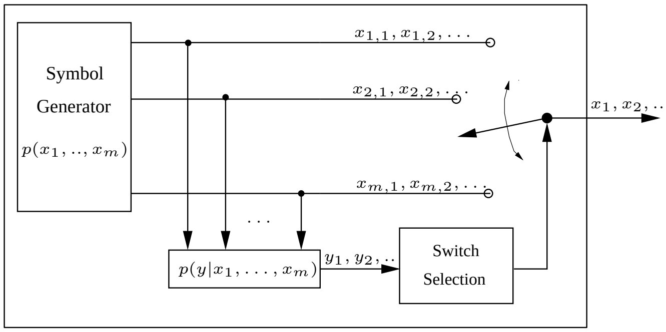

Fig. 2. A simple model for an active source that observes the environment noisily.

is based on purpose of the system it is a part of. From the perspective of the coding system, the output of the active source could be carefully modelled based on this objective and the rules implemented by the agent controlling the source. However, actually formulating such a model may be difficult, and even after doing so, the rate-distortion function for such a source may be unknown in a single-letter or otherwise computable form. For these reasons, we focus on several generic models for an active source: adversarial (worst-case), helpful (joint optimization of coding and active vision systems) and random (agnostic). The idea is that, regardless of the *modus operandi* of the active source, the rate required to compress its output will lie somewhere between the worst-case and joint optimization scenarios. Our results along with the examples show that it can be very useful to jointly optimize the coding system with the active vision source.

We note that these issues are not confined to vision, but can arise in other applications. For example, consider a field of sensors taking measurements and an autonomously moving agent that collects the samples and decides its path based on samples it has already taken from the field. If there is correlation in the sensor field, some degree of noncausality may be induced on the active agent.

The paper is organized as follows. In Section II, we introduce notation, the model of an active source and review a few relevant results. In Section III, we determine the rate-distortion function for adversarial active sources. In Section IV, we study 'helpful' active sources and determine the rate-distortion function in a special case. Some examples illustrating the gaps between helpful and adversarial active sources are give in Section V. We conclude in Section VI.

II. Problem Setup and Literature Review

A. Notation and Model

In this section, we set up the active vision source as a special kind of arbitrarily varying source. Let $\mathcal{X}$ and $\widehat{\mathcal{X}}$ be finite source and reconstruction alphabets respectively. We will let $\mathbf{x}^n$ denote a length n vector $(x_1,\ldots,x_n)\in\mathcal{X}^n$. Similarly, $\widehat{\mathbf{x}}^n$ denotes a vector from $\widehat{\mathcal{X}}^n$. As needed, $\mathbf{x}^k$ will denote the first k letters in a vector $\mathbf{x}^n$. Let $d:\mathcal{X}\times\widehat{\mathcal{X}}\to[0,d^*]$ be a distortion measure for some $d^*<\infty$. For $n\ge 1$, let $d_n:\mathcal{X}^n\times\widehat{\mathcal{X}}^n\to[0,d^*]$ be the average distortion $d_n(\mathbf{x}^n,\widehat{\mathbf{x}}^n)=\frac{1}{n}\sum_{k=1}^n d(x_k,\widehat{x}_k)$. Let $\mathcal{P}(\mathcal{X})$ denote the set of distributions (probability mass functions) on $\mathcal{X}$, and let $\mathcal{P}_n(\mathcal{X})$ denote the set of types for length n vectors in $\mathcal{X}$. Let $\mathcal{W}$ be the set of probability transition matrices (channels) from $\mathcal{X}$ to $\widehat{\mathcal{X}}$. For a $p\in\mathcal{P}(\mathcal{X})$, we let $D_{\min}(p)=\sum_{x\in\mathcal{X}}p(x)\min_{\widehat{x}\in\widehat{\mathcal{X}}}d(x,\widehat{x})$. $D_{\min}(p)$ is the minimum distortion achievable on average for an IID source with distribution p. The rate-distortion function for a distribution p at distortion[2] $D\ge D_{\min}(p)$ is

$$R(p,D)=\min_{W\in\mathcal{W}(p,D)} I(p,W), \tag{1}$$

where

$$\mathcal{W}(p,D)=\left\{W\in\mathcal{W}:\sum_{x,\widehat{x}}p(x)W(\widehat{x}|x)d(x,\widehat{x})\le D\right\} \tag{2}$$

and $I(p,W)$ is the mutual information

$$I(p,W)=\sum_{x,\widehat{x}}p(x)W(\widehat{x}|x)\log_2\frac{W(\widehat{x}|x)}{\sum_{x'\in\mathcal{X}}p(x')W(\widehat{x}|x')}. \tag{3}$$

Let $\mathcal{B}=\{\widehat{\mathbf{x}}^n(1),\ldots\widehat{\mathbf{x}}^n(M)\}$ be a reconstruction codebook with M elements from $\widehat{\mathcal{X}}^n$. Let $d(\mathbf{x}^n;\mathcal{B})=\min_{\widehat{\mathbf{x}}^n\in\mathcal{B}}d_n(\mathbf{x}^n,\widehat{\mathbf{x}}^n)$ be the distortion between $\mathbf{x}^n$ and the codeword from $\mathcal{B}$ that best represents it.

Figures 2 and 3 show two equivalent models for an active source. At each time, the environment is modelled by m correlated random variables $x_1,\ldots,x_m$ that each take values in $\mathcal{X}$. The m random variables are referred to as subsources, and are IID across time. Let $x_{l,k}$ denote the output of the l^{th} subsource at time k. We assume there is a known probability distribution on the subsources, so that for any time k, $P(x_{1,k}=x_1,\ldots,x_{m,k}=x_m)=p(x_1,\ldots,x_m)$.

The output of the active source at time k is determined by a switch position, $s_k\in\{1,\ldots,m\}$. With a slight abuse of notation, let x_k be the output of the active source at time k, so that $x_k=x_{s_k,k}$. The agent, called a 'switcher', chooses the switch positions based on observations y_k of the environment seen through a discrete memoryless channel $p(y_k|x_{1,k},\ldots,x_{m,k})$ in Figure 2. When all distributions are known, this model is equivalent (by Bayes' rule) to the one shown in Figure 3 where the switcher has access to a memoryless state t_k at each time, and the subsources output the $x_{l,k}$ with conditional distribution $p(x_1,\ldots,x_m|t)$.

Let $\mathcal{T}$ be the finite set of states and let $\{p_l(\cdot|t)\}_{l=1}^m\subset\mathcal{P}(\mathcal{X})$ be the marginals of the conditional distribution $p(x_1,\ldots,x_m|t)$. For each $t\in\mathcal{T}$, we let $\overline{\mathcal{G}}(t)$ denote the convex hull of $\{p_1(\cdot|t),\ldots,p_m(\cdot|t)\}$. We let $\alpha(t)$ denote the distribution on the state t, and note that $t_1,t_2,\ldots$ is assumed to be IID with distribution α.

In this paper, we focus on adversarial and helpful models for the active source. For $n\ge 1$, $D\ge 0$, we let

$$M^{adv}(n,D)=\min\left\{|\mathcal{B}|:\begin{array}{c}\mathcal{B}\subset\widehat{\mathcal{X}}^n,\ \mathbb{E}[d_n(\mathbf{x}^n;\mathcal{B})]\le D\\ \text{for all allowable}\\ \text{switcher strategies}\end{array}\right\}$$

[2] By convention, $R(p,D)=\infty$ for $D<D_{\min}(p)$.

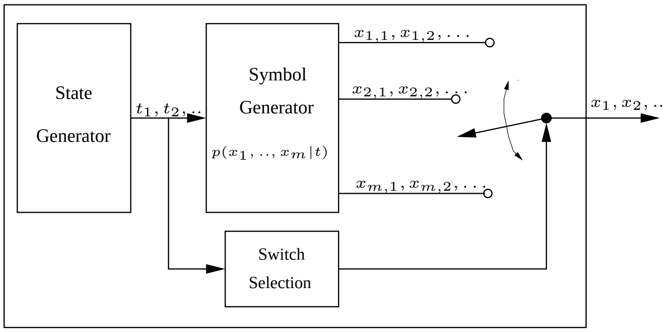

Fig. 3. A model equivalent to that of Figure 2, where the active source observes a state for the environment.

and[3]

$$M^{help}(n,D)=\min\left\{|\mathcal{B}|:\begin{array}{c}\mathcal{B}\subset\widehat{\mathcal{X}}^n,\ \mathbb{E}[d_n(\mathbf{x}^n;\mathcal{B})]\le D\\ \text{for some allowable}\\ \text{switcher strategy}\end{array}\right\}.$$

The expectations in the above definitions depend on the switcher strategy and whether the agent is causal or noncausal, but in general, $\mathbb{E}[d_n(\mathbf{x}^n;\mathcal{B})]=\sum_{\mathbf{x}^n}P(\mathbf{x}^n)d(\mathbf{x}^n;\mathcal{B})$ where

$$P(\mathbf{x}^n)=\sum_{\mathbf{s}^n,\mathbf{t}^n}\left[\prod_{k=1}^n\alpha(t_k)\right]P(\mathbf{s}^n,\mathbf{x}^n|\mathbf{t}^n). \tag{4}$$

Above, $P(\mathbf{s}^n,\mathbf{x}^n|\mathbf{t}^n)$ represents the switcher's allowable strategy. If the agent is restricted to use memoryless rules based only on the current state, for example, $P(\mathbf{s}^n,\mathbf{x}^n|\mathbf{t}^n)=\prod_{k=1}^n P_k(s_k|t_k)p_{s_k}(x_k|t_k)$.

We define three classes of switchers. First, a strictly causal switcher is allowed to choose the switch position s_k at time k as a function of $\mathbf{x}_1^{k-1},\ldots,\mathbf{x}_m^{k-1}$ and $\mathbf{t}^{k-1}$. A causal switcher can choose s_k as a function of $\mathbf{x}_1^{k-1},\ldots,\mathbf{x}_m^{k-1}$ and $\mathbf{t}^k$. Finally, a noncausal switcher can choose s_k as a function $\mathbf{x}_1^{k-1},\ldots,\mathbf{x}_m^{k-1}$ and $\mathbf{t}^n$. We refer to the causal switcher as having 1-step lookahead (as compared to the strictly causal switcher) and the noncausal switcher as having full lookahead.

The measure of performance is the asymptotic rate needed to compress the output of the active source to within a distortion D according to the model. Hence, we define

$$R^{adv}(D)=\limsup_{n\to\infty}\frac{1}{n}\log_2 M^{adv}(n,D) \tag{5}$$

and

$$R^{help}(D)=\limsup_{n\to\infty}\frac{1}{n}\log_2 M^{help}(n,D). \tag{6}$$

B. Literature Review

Several special cases of the model in Section II-A have already been studied in the literature. Most obviously, if $m=1$, switching is meaningless and we have the classic rate-distortion problem for an IID source with distribution p. Shannon [3] showed that for an IID source $R^{adv}(D)=R^{help}(D)=R(p,D)$.

[3]By convention, if no such $\mathcal{B}$ exists, $M(n,D)=\infty$ in these definitions.

For the strictly causal switcher, when the state is independent (i.e. the state is useless to the agent for inferring anything about the environment) of the subsource outputs, Berger [4] showed that

$$R^{adv}(D)=\max_{p\in\overline{\mathcal{G}}}R(p,D), \tag{7}$$

where $\overline{\mathcal{G}}$ is the convex hull of the distributions on the subsources. In the context of active vision, a strictly causal adversary pointing a camera is intuitively no more threatening than a robot randomly pointing the camera when the scene being captured is memoryless.

In [5], causal and noncausal adversaries are considered when the state reveals the realizations of the m subsources noiselessly to the switcher. It was shown that if the switcher has 1-step lookahead (causal) or full lookahead (noncausal),

$$R^{adv}(D)=\max_{p\in\mathcal{C}}R(p,D), \tag{8}$$

where

$$\mathcal{C}=\left\{p\in\mathcal{P}(\mathcal{X}):\begin{array}{c}p(\cdot)=\sum_{\mathcal{V}\subset\mathcal{X},1\le|\mathcal{V}|\le m}\alpha(\mathcal{V})f(\cdot|\mathcal{V})\\ f(\cdot|\mathcal{V})\in\mathcal{P}(\mathcal{V})\end{array}\right\}, \tag{9}$$

$\alpha(\mathcal{V})$ is the probability that $\{x_{1,k},\ldots,x_{m,k}\}=\mathcal{V}$ and $\mathcal{P}(\mathcal{V})$ is the set of distributions on $\mathcal{V}$. The fact that causal and noncausal adversaries have the same rate-distortion function is a by-product of the assumption that the environment is memoryless. It was also shown by example that $R^{adv}(D)$ can increase when the switcher is allowed to be causal or noncausal as opposed to strictly causal.

III. Adversarial Active Sources

For active sources modelled as in Section II-A, we have the following theorem.

Theorem 1: For both causal and noncausal active sources,

$$R^{adv}(D)=\max_{p\in\mathcal{D}}R(p,D), \tag{10}$$

where

$$\mathcal{D}=\left\{p\in\mathcal{P}(\mathcal{X}):\begin{array}{c}p(\cdot)=\sum_{t\in\mathcal{T}}\alpha(t)f(\cdot|t),\\ f(\cdot|t)\in\overline{\mathcal{G}}(t)\ \forall\ t\in\mathcal{T}\end{array}\right\}.$$

Proof: (Outline) For the complete proof, see [6]. First, for the converse part, let $p\in\mathcal{D}$ be a distribution that achieves $R^{adv}(D)$ in equation (10). Upon observing a state t_k at time k, the switcher can set the switch position s_k randomly according to the convex combination that yields $f(\cdot|t_k)$. The switcher needs only 1-step lookahead to enact this strategy, and hence it is also allowable if the switcher has full lookahead. To the coding system, the output of the active source looks IID with distribution p. Therefore, to code such a source to within distortion D, it is required that $R^{adv}(D)\ge\max_{p\in\mathcal{D}}R(p,D)$.

For the direct part, we use the type covering lemma of [4] and [7]. The type covering lemma states that if $p\in\mathcal{P}_n(\mathcal{X})$ for large enough n, all vectors with type p can be covered to within distortion D with at most $\exp_2(n(R(p,D)+\epsilon))$ codewords. Since the number of types only grows polynomially with n, we can take a union of codebooks over types within $\mathcal{D}$ without asymptotically affecting the rate. The rate of this new codebook is dominated by the largest individual codebook, the rate of which can be made arbitrarily close to $\max_{p\in\mathcal{D}} R(p, D)$ for large enough n. Hence, we will have covered all vectors with types in $\mathcal{D}$, so we need only show that the type of a vector output by the active source must lie within (or very near) $\mathcal{D}$ with high probability. This can be done by means of a martingale argument that can be found in Lemma 2.1 of Appendix II in [6]. ■

Note that when the agent can observe the environment noiselessly, the set $\mathcal{D}$ equates directly with the set $\mathcal{C}$ of equation (9). Computing $R^{adv}(D)$ may be difficult because the IID rate-distortion function $R(p, D)$ is generally not concave in the distribution p. In [6], we give a 'brute-force' algorithm to find $R^{adv}(D)$ to within some precision $\epsilon > 0$.

IV. Helpful Active Sources

For helpful active sources with 1-step lookahead, we have the following immediate lemma.

Lemma 1: If the active source has 1-step lookahead,

$$R^{help}(D) \leq \min_{p\in\mathcal{D}} R(p, D). \tag{11}$$

Proof: Again, if we let p be a distribution that achieves the minimization in (11), the switcher can simulate the distribution $f(\cdot|t)$ upon observing a state t. The resulting output of the active source looks to the coding system like an IID source with distribution p. ■

In the special case that the source has full lookahead and the state is exactly the output of the m subsources, we can characterize the rate-distortion function exactly as the IID rate-distortion function for an associated source. Let

$$\mathcal{X}^* = \{\mathcal{V} \subseteq \mathcal{X} : 1 \leq |\mathcal{V}| \leq m\} \tag{12}$$

and define a new distortion measure $\rho : \mathcal{X}^* \times \widehat{\mathcal{X}} \to [0, d^*]$ by

$$\rho(\mathcal{V}, \widehat{x}) = \min_{x\in\mathcal{X}} d(x, \widehat{x}). \tag{13}$$

Let $\mathcal{V}_k = \{x_{1,k}, \ldots, x_{m,k}\}$ be the sequence of IID 'observed sets' with distribution $\alpha(\mathcal{V})$ and let $R^*(\alpha, D)$ be the rate-distortion function for the distribution α with respect to distortion measure ρ.

Theorem 2: If the active source observes the subsource realizations noiselessly and has full lookahead,

$$R^{help}(D) = R^*(\alpha, D). \tag{14}$$

Proof: For any given codebook $\mathcal{B}$, the helpful switcher will try to output letters $x_1, \ldots, x_n$ from $\mathcal{V}_1, \ldots, \mathcal{V}_n$ respectively so that $d_n(\mathbf{x}^n; \mathcal{B})$ is minimized. For any fixed $\widehat{\mathbf{x}}^n$ in the codebook, the minimal distortion $\mathbf{x}^n$ sequence that can be output by the active source is such that

$$d_n(\mathbf{x}^n, \widehat{\mathbf{x}}^n) = \frac{1}{n}\sum_{k=1}^{n} \rho(\mathcal{V}_k, \widehat{x}^k). \tag{15}$$

Hence, we have by proper selection of the switch positions,

$$d_n(\mathbf{x}^n; \mathcal{B}) = \min_{\widehat{\mathbf{x}}^n\in\mathcal{B}} \frac{1}{n}\sum_{k=1}^{n} \rho_n(\mathcal{V}_k, \widehat{x}_k). \tag{16}$$

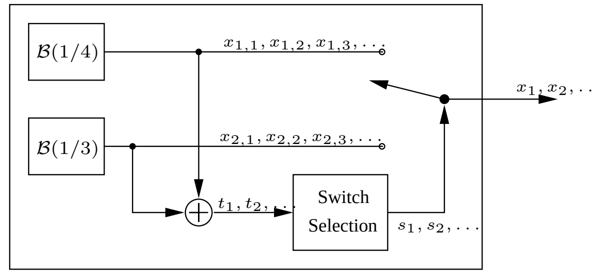

Fig. 4. Two Bernoulli subsources, the agent controlling the active source observes the mod-2 sum of the two subsource outputs.

The problem of covering the $\mathbf{x}^n$ vector to within distortion D with respect to $d(\cdot,\cdot)$ is exactly the same as covering the sets $\mathcal{V}_1, \ldots, \mathcal{V}_n$ to within distortion D with respect to $\rho(\cdot,\cdot)$. ■

The strategy that achieves a rate of $R^*(\alpha, D)$ requires the active source to process the entire block and then select the switch positions. This is not possible with 1-step lookahead (or if the active sensor does not have the computational capabilities to process a long block of measurements), so for causal agents, we have the following corollary.

Corollary 1: For causal switchers,

$$R^*(\alpha, D) \leq R^{help}(D) \leq \min_{p\in\mathcal{D}} R(p, D). \tag{17}$$

In the next section, we give an example showing that $R^*(\alpha, D) < \min_{p\in\mathcal{D}} R(p, D) < \max_{p\in\mathcal{D}} R(p, D)$ in general.

V. Examples

To simply illustrate the results, we consider examples using binary alphabets $\mathcal{X} = \widehat{\mathcal{X}} = \{0, 1\}$ and Hamming distortion, $d(x, \widehat{x}) = 1(x \neq \widehat{x})$. Recall that for an IID binary (Bernoulli) source with a probability of 1 equal to $p \in [0, 1/2]$,

$$R((1-p, p), D) = \begin{cases} h_b(p) - h_b(D) & D \in [0, p] \\ 0 & D > p \end{cases}, \tag{18}$$

where $h_b(p) = -p\log_2 p - (1-p)\log_2(1-p)$ is the binary entropy function. The examples consider $m = 2$ independent Bernoulli subsources with a probability of 1 equal to $1/4$ and $1/3$ for the first and second subsource respectively. For the strictly causal adversary, the rate-distortion function is $R^{adv}(D) = h_b(1/3) - h_b(D)$ for $D \in [0, 1/3]$, as we know from [4]. We know from [5] that if the adversary is causal or noncausal and the observations are the noiseless realizations of the subsources, $R^{adv}(D) = 1 - h_b(D)$ for $D \in [0, 1/2]$. Figure 4 shows an example where the mod-2 sum of the subsources is observed by the switcher. By evaluating Theorem 1, we see that $R^{adv}(D) = h_b(1/3) - h_b(D)$ for $D \in [0, 1/3]$. Therefore, observation of the mod-2 sum of the subsources does not allow the adversarially active source to increase the rate-distortion function above that of the strictly causal adversary.

Figure 5 shows an example where the agent observes only the second subsource, but not the first. Again using Theorem 1, we see that for the causal or noncausal active

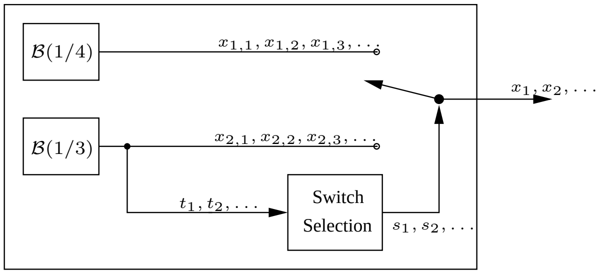

Fig. 5. Two Bernoulli subsources, the second of which is observed noiselessly by the agent controlling the active source.

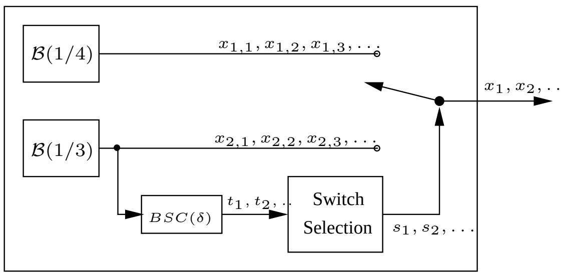

Fig. 6. Two Bernoulli subsources with parameters $1/4$ and $1/3$. The $1/3$ source is observed through a binary symmetric channel with crossover probability $\delta \in [0, 1/2]$.

source, $R^{adv}(D) = 1 - h_b(D)$. In this example, the rate-distortion function is the same as if the agent had observed both subsources.

Figure 6 shows an example where the second subsource is observed by the switcher through a binary symmetric channel with crossover probability $\delta \in [0, 1/2]$. For the causal or noncausal active source, if $\delta \in [0, 2/5]$, it can be shown that

$$R^{adv}(D) = h_b\left(\frac{1}{2} - \frac{5\delta}{12}\right) - h_b(D), D \in \left[0, \frac{1}{2} - \frac{5}{12}\delta\right]. \tag{19}$$

For $\delta > 2/5$, $R^{adv}(D) = h_b(1/3) - h_b(D)$ for $D \in [0, 1/3]$ as the observations are too noisy to increase the rate-distortion function over the strictly causal case.

Now, if the active source in these examples observes both subsources noiselessly, with 1-step lookahead, $R^{help}(D) \leq h_b(1/12) - h_b(D)$ for $D \in [0, 1/12]$ by Lemma 1. If the active source has full lookahead, we can compute $R^{help}(D)$ from Theorem 2 using the Arimoto-Blahut algorithm [8]. All these rate-distortion functions are plotted for comparison in Figure 7. The plot shows that there is a large difference between the rate-distortion functions for adversarial active sources and their helpful counterparts. When the agent is allowed to see the subsources noiseless for these examples, we have the strict inequality $R^*(\alpha, D) < \min_{p \in \mathcal{D}} R(p, D) < \max_{p \in \mathcal{D}} R(p, D)$.

VI. Concluding Remarks

In order to truly understand lossy compression of active sources, there are several major issues that further need to be dealt with. Most obviously, subsources with memory should be studied. When the subsources have memory, causality will

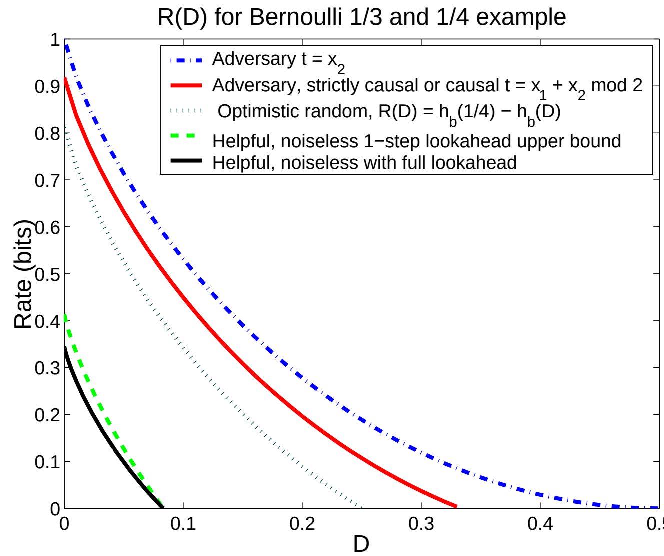

Fig. 7. Comparison of rate-distortion functions for various examples. There is a large gap between helpful active sources and adversarial active sources.

likely affect the rate distortion function in a much more gradual way.

Another important consideration is whether the active vision source is part of a closed-loop control system or not. If it is, delay becomes a critical issue and it may be worthwhile to study the problem within the framework of causal source coding [9].

Finally, we have shown that large savings in rate can be had by jointly optimizing the coding system with the active vision system. To some extent, these savings are artificial as they completely disregard the objective of the active vision system. However, the question of how to jointly optimize the two systems while maintaining some adequate level of performance within each is a worthwhile one and could lead to an interesting tradeoff involving mismatched distortion measures.

References

[1] R. Bajcsy, "Active perception," *Proceedings of the IEEE*, vol. 76, no. 8, pp. 966–1005, Aug. 1988.
[2] A. Chebira, P. Dragotti, L. Sbaiz, and M. Vetterli, "Sampling and interpolation of the plenoptic function," in *Proc.of IEEE International Conference on Image Processing*, Barcelona, Spain, Sept. 2003.
[3] C. Shannon, "Coding theorems for a discrete source with a fidelity criterion," in *IRE Natl. Conv. Rec.*, 1959, pp. 142–163.
[4] T. Berger, "The source coding game," *IEEE Trans. Inform. Theory*, vol. 17, pp. 71–76, Jan. 1971.
[5] H. Palaiyanur, C. Chang, and A. Sahai, "The source coding game with a cheating switcher," in *Proc. Int. Symp. Inform. Theory*, Nice, France, June 2007.
[6] ——, "The source coding game with a cheating switcher." [Online]. Available: http://arxiv.org/abs/0712.2870
[7] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. New York, NY: Academic Press, 1997.
[8] T. Cover and J. Thomas, *Elements of Information Theory*. New York, NY: John Wiley and Sons, 1991.
[9] D. Neuhoff and R. K. Gilbert, "Causal source codes," *IEEE Trans. Inform. Theory*, vol. 28, pp. 701–713, Sept. 1982.