



**MATEMATICKO-FYZIKÁLNÍ  
FAKULTA**  
Univerzita Karlova

## **BAKALÁŘSKÁ PRÁCE**

Viktor Němeček

# **Percepční učení a Ideální Bayesovský pozorovatel při zrakovém vyhledávání**

Katedra softwaru a výuky informatiky

Vedoucí bakalářské práce: Mgr. Filip Děchtěrenko, Ph.D.

Studijní program: Informatika

Studijní obor: Obecná informatika

Praha 2018

Prohlašuji, že jsem tuto bakalářskou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V ..... dne .....

Podpis autora

Poděkování.

Název práce: Percepční učení a Ideální Bayesovský pozorovatel při zrakovém vyhledávání

Autor: Viktor Němeček

Katedra: Katedra softwaru a výuky informatiky

Vedoucí bakalářské práce: Mgr. Filip Děchtěrenko, Ph.D., Katedra softwaru a výuky informatiky

Abstrakt: Abstrakt.

Klíčová slova: klíčová slova

Title: Perceptual learning and Ideal Bayesian observer in visual search task

Author: Viktor Němeček

Department: Department of Software and Computer Science Education

Supervisor: Mgr. Filip Děchtěrenko, Ph.D., Department of Software and Computer Science Education

Abstract: Abstract.

Keywords: key words

# Obsah

<b>Úvod</b>	<b>2</b>
<b>1 Základní pojmy</b>	<b>3</b>
1.1 Šum . . . . .	3
1.1.1 Barva šumu . . . . .	3
1.2 Gabor patch . . . . .	3
1.2.1 Definice . . . . .	3
1.2.2 Použití . . . . .	5
1.3 Úvod do teorie detekce signálu . . . . .	5
1.3.1 Senzitivita . . . . .	6
1.4 Entropie . . . . .	8
1.5 Modely pozorovatele . . . . .	8
1.5.1 $d'$ mapa . . . . .	9
1.5.2 Modely chování pozorovatele . . . . .	9
<b>2 Metody</b>	<b>12</b>
2.1 Účastníci . . . . .	12
2.2 Nástroje a stimuly . . . . .	12
2.3 Procedura . . . . .	12
2.4 Limitace . . . . .	13
<b>3 Měření</b>	<b>15</b>
3.1 Metodika . . . . .	15
3.2 Výsledky . . . . .	15
<b>4 Implementace</b>	<b>16</b>
<b>Závěr</b>	<b>17</b>
<b>Seznam použité literatury</b>	<b>18</b>
<b>Seznam obrázků</b>	<b>20</b>
<b>Seznam použitých zkratk</b>	<b>21</b>
<b>A Přílohy</b>	<b>22</b>
A.1 První příloha . . . . .	22

# Úvod

Následuje několik ukázkových kapitol, které doporučují, jak by se měla bakalářská práce sázet. Primárně popisují použití T<sub>E</sub>Xové šablony, ale obecné rady poslouží dobře i uživatelům jiných systémů.

# 1. Základní pojmy

## 1.1 Šum

V teorii detekce signálu šumem nazýváme jakoukoli nechtěnou (a typicky neznámou) modifikaci signálu.

### 1.1.1 Barva šumu

U aditivního<sup>1</sup> šumu můžeme měřit intenzitu šumu na různých frekvencích. Tu změříme tak, že na šum aplikujeme Fourierovu transformaci

Šumu se říká bílý šum, pokud by světlo, které by mělo stejnou distribuci intenzity napříč frekvencemi, jako daný šum (který ale vůbec nemusí být světelný), bylo bílé. Podobně známe například ještě růžový, červený či modrý šum.

Intenzita  $p$  všech zmíněných šumů na  $v$  na dané frekvenci lze vyjádřit jako  $p = 1/f^\beta$ , kde hodnota  $\beta$  je  $-1$  pro modrý šum,  $0$  pro bílý,  $1$  pro růžový a  $2$  pro hnědý. Proto se růžový šum někdy též označuje jako  $1/f$  šum. Pro ostatní barvy šumu není podobné označení běžné.

Pravděpodobnostní rozdělení jednotlivých složek Fourierovy transformace ale není definicí barvy šumu dáno. Pokud je rozdělení normální<sup>2</sup>, řekneme, že se jedná o Gaussovský šum. **Tyhle obecný kecy moc nikde na jednom místě krom Wikipedie nejsou – něco málo je tady <http://www.atis.org/glossary/> – Nějaká americká telekomunikační norma.**

V této práci se budeme zabývat vizuálním šumem, tedy šumem, kde místo obvykle používané časové souřadnice použijeme dvě souřadnice prostorové, a měřenou hodnotou bude jas.

## 1.2 Gabor patch

Gabor filter (v českých textech někdy označovaný jako Gaborova vlnka) je lineární filtr používaný ve zpracování obrazu, chceme-li detekovat signál mající danou frekvenci a směr, který se vyskytuje kolem daného bodu.

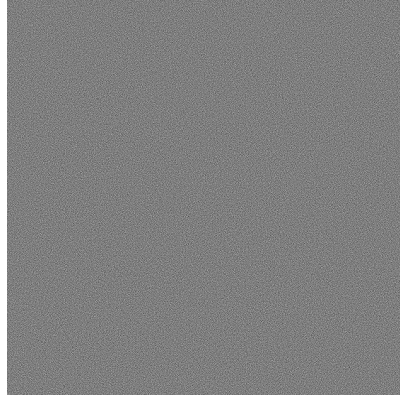
### 1.2.1 Definice

Hodnotu filtru v daném bodě spočítáme jako součin dvou funkcí. První z nich je vždy sinus či cosinus (někdy uváděné v podobě komplexní exponenciály, pokud potřebujeme i reálnou, i imaginární složku). Jeho parametry určují, jaké vlastnosti má mít signál, který chceme detekovat. Druhé funkci říkáme obálka, a určuje, na jakém okolí daného bodu signál zkoumáme.

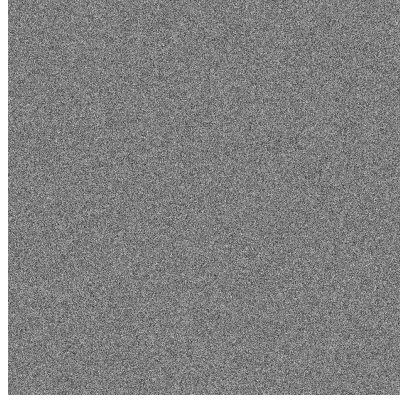
---

<sup>1</sup>Šumu říkáme aditivní, pokud se jeho hodnoty přičítají k hodnotě signálu. Dále existuje například ještě šum multiplikativní či fázový (šum, který se projevuje krátkodobým fázovým posunem signálu).

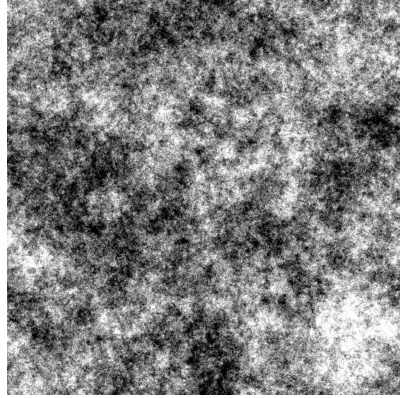
<sup>2</sup>Nezáleží na tom, zda má normální rozdělení šum sám nebo jeho Fourierův obraz – tyto dvě vlastnosti jsou ekvivalentní.



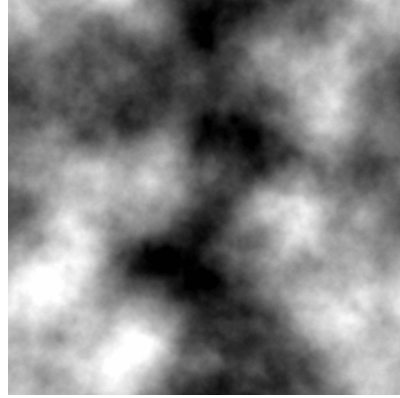
(a) Modrý šum



(b) Bílý šum



(c) Růžový šum



(d) Červený, někdy též Brownův šum

Obrázek 1.1: Ukázky různých šumů.

Funkce tedy vypadá jako

$$g(x,y) = \sin\left(2\pi\frac{x'}{\lambda} + \phi\right) * \text{obálka}(x',y'),$$

kde vektor  $(x',y')^T$  je vektor  $(x,y)^T$  otočený o úhel, který svírá osa  $x$  se směrem, podél něž chceme měřit signál (tento úhel budeme značit  $\Theta$ ), a posunutý do bodu, v němž chceme měřit signál,  $\lambda$  je frekvence signálu, který hledáme, a  $\phi$  je fázový posun (Fogel a Sagi, 1989).

Jako obálka se používá dvojrozměrná Gaussova funkce, raised cosine, nebo prostá lineární funkce vzdálenosti.

Gaussovu funkci vyjádříme jako

$$\text{obálka}(x,y) = \exp\left(-\frac{x'^2 + y'^2}{2\rho}\right),$$

kde  $\rho$  je směrodatná odchylka Gaussovy křivky. Její výhodou je, že chování Gabor filtru, jehož obálku tvoří Gaussova funkce, je nejlépe popsané. Raised cosine vyjádříme jako

$$\text{obálka}(x,y) = \begin{cases} \frac{\cos(\pi\sqrt{x'^2+y'^2}/r)+1}{2} & \text{pro } \sqrt{x'^2+y'^2} \leq r, \\ 0 & \text{jinak,} \end{cases}$$



kde  $r$  je poloměr oblasti, v níž chceme signál detekovat. Výhodou raised cosine oproti Gaussově funkci je, že ve vzdálenosti alespoň  $r$  od středu filtru jeho hodnota nabývá nuly. Při výpočtech tedy stačí počítat s malou oblastí kolem středu (kdežto při použití Gaussovy funkce je nutné počítat s celým obrazem). Výhodou oproti lineární funkci vzdálenosti je, že raised cosine se pro většinu aplikací chová dostatečně podobně, jako Gaussova funkce.

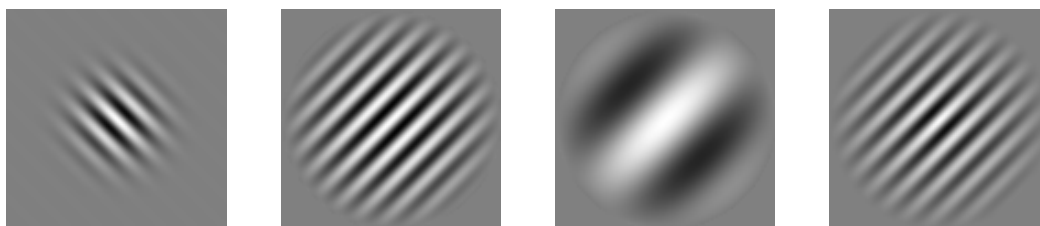
### 1.2.2 Použití

Chceme-li detekovat signál ve vizuálním šumu, spočítáme hodnotu

$$s = \sum g(x,y) * n[x,y],$$

kde  $n$  je šum a sumu bereme přes všechny body  $(x,y)$ , v nichž jsme naměřili hodnoty šumu. Je-li hodnota  $s$  blízko nuly, signál v daném místě není přítomen, nebo je přítomen s jinými parametry. Vysoké hodnoty značí, že signál pravděpodobně přítomen je, hluboce záporné značí, že signál je přítomen, ovšem s fází posunutou  $\pi$ .

Gabor filter ale můžeme používat i k samotné tvorbě signálu. Chceme-li vytvořit v nějakém bodě signál, můžeme spočítat Gabor filter, jako bychom chtěli detekovat signál s právě takovými parametry, jaké má mít tvořený signál, a potom ho sečíst se šumem. Takto vytvořenému signálu budeme říkat Gabor patch.



Obrázek 1.2: Ukázky několika Gabor patchů. Všechny gabor patche jsou 100 pixelů široké i vysoké. Levý patch má  $\Theta = 1/4\pi$ , ostatní mají  $\Theta = -1/4\pi$ , levý má jako obálku Gaussovu funkci, prostřední dva raised cosine, pravý lineární funkci vzdálenosti, první, druhý a čtvrtý mají frekvenci (v cyklech na pixel) 0.1, třetí 0.02.

## 1.3 Úvod do teorie detekce signálu

Obsah této sekce je volně převzatý od Jozefa Amemoriho (2016).

Teorie detekce signálu řeší problém rozlišení dvou signálů, případně signálu a šumu. V této práci se budem zabývat problémem, kdy je potřeba rozhodnout, zda se v dané ploše nachází signál, či ne. Této úloze budeme říkat A/N problem, nebo pouze A/N.

V kontextu této úlohy můžou po odpovědi pozorovatele nastat čtyři možné situace:

	Signál je přítomen	Signál není přítomen
Pozorovatel odpoví kladně	<i>Hit</i>	<i>False alarm</i>
Pozorovatel odpoví záporně	<i>Miss</i>	<i>Correct rejection</i>

V této úloze se pozorovatel chová tak, že si zvolí kritérium (někdy též práh odpovědi), které udává, jaká musí být pravděpodobnost, že signál je přítomen, aby pozorovatel odpověděl kladně (Green a Swets, 1988, kapitola 1.7).

Poté pozorovatel provede pozorování, z něž získá informaci  $s$ . Poté spočítá hodnotu rozhodovací proměnné, tedy spočítá, jaká je pravděpodobnost, že je signál přítomen. Tato pravděpodobnost se z Bayesovy věty spočítá jako

$$P[\text{Signál je přítomen} | \text{Pozorování dalo informaci } s] = \\ = \frac{P[\text{Pozorování dalo informaci } s | \text{Signál je přítomen}] * P[\text{Signál je přítomen}]}{P[\text{Pozorování dalo informaci } s]}$$

Poté porovná tuto rozhodovací proměnnou s kritériem a podle výsledku tohoto porovnání odpoví.

Odsud je zřejmé, že zvyšováním kritéria zvedáme pravděpodobnost, že nastane miss, ale snižujeme pravděpodobnost false alarmu.

V souvislosti s A/N úlohou je ale ještě nutné zavést následující pojmy:

**Definice 1.** *Mějme pozorovatele  $p$  v úloze A/N. Potom jeho hit rate  $H_p$  definujeme jako*

$$P[\text{hit} | \text{Signál je přítomen}]$$

*a jeho False alarm rate  $F_p$  jako*

$$P[\text{False alarm} | \text{Signál není přítomen}].$$

### 1.3.1 Senzitivita

Jedním s dílčích problémů je též určení, jakou rozlišovací schopnost má daný pozorovatel na daný signál. Z předchozího odstavce by mělo být vidět, že kvantifikovat citlivost pouze pomocí hit rate není vhodné – například pozorovatel, který vždy odpoví „ano“ by měl v tomto případě optimální citlivost, což jistě nechceme. Taktéž není vhodné charakterizovat pozorovatele pomocí poměru hitů a false alarmů, optimalizace takového parametru by též vedla k scestnému nastavení kritéria.<sup>3</sup>

Další možností by bylo charakterizovat pozorovatele jeho poměrem správných a špatných odpovědí. To ale též není vhodné, protože potom by jeho citlivost nezávisela jen na jeho vlastnostech, ale v případě, kdy by pravděpodobnost hitu a pravděpodobnost correct rejection nebyla stejná, i na poměru počtu situací, kdy je signál přítomen, v naší sadě testů.

Při detekci signálu rozeznáváme dva druhy šumu. Šum vnitřní a šum vnější (Heeger, 1997). Vnější šum je zkreslení dat, které se stane ještě mimo pozorovatele. Vnitřní šum je naopak šum, který je způsoben přímo algoritmem, který používá pozorovatel, aby dekodoval, zda signál je přítomen, či není.

Pro lepší porozumnění představíme případ z praxe. Necht je pozorovatel člověk, který není příliš pozorný a chystá se přejít silnici. Jeho (podvědomým) pozorováním je informace, která dorazí ddo jeho mozku z periférií jeho sítnic, a

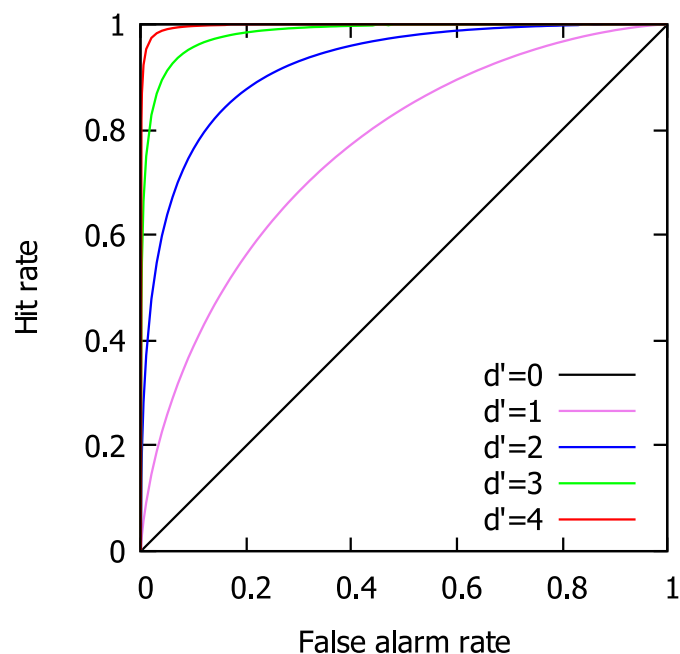
<sup>3</sup>V typickém případě, který je popsán v následujících třech odstavcích, se pro kritérium jdoucí k nekonečnu tento poměr též blíží k nekonečnu. Dobrý pozorovatel by tedy téměř vždy odpovídal „ne“, optimální pozorovatel by neexistoval.

signál, jehož přítomnost zkoumá, je, zda přijíždí auto. V takovém případě se v mozku spustí algoritmus, který převede pozorování na číslo  $X$ . Poznamenejme, že v mozku je hodnota tohoto čísla dobře kvantifikovaná, i když je mimo záběr této práce popisovat, co přesně z fyzikálního hlediska tato hodnota znamená. Pro porozumnění stačí představa, že se jedná o frekvenci určitého druhu impulsu, které si mezi sebou neurony posílají. Mozek ví, jakou distribuci má náhodná veličina  $X$  v případě, že se auto blíží, a jakou má v případě, že nikoliv. Podle konkrétní hodnoty  $X$  kterou naměřil, se rozhodne, jestli auto přítomno není, a je tedy bezpečné silnici přejít, nebo zda je pravděpodobnost přijíždějícího auta dost vysoká na to, aby mělo smysl zvednout hlavu a podívat se příslušným směrem znovu s cílem získat lepší pozorování.

Praxe ukazuje, že v těchto situacích je rozdělení veličiny  $X$  často dobře aproximováno normálním rozdělením, a to jak v případě, že signál přítomen je, tak v případě, že přítomen není. Obě tato normální rozdělení navíc mají v tomto typickém případě stejný rozptyl a liší se jen svou střední hodnotou (Swets, 1996). Potom je přirozená otázka ptát se, jaký je rozdíl mezi jejich středními hodnotami (či lépe rozdíl mezi jejich středními hodnotami vydělený směrodatnou odchylkou).

Právě tuto hodnotu označíme jako hodnotu  $d'$ . Protože ale ne vždy je rozdělení normální, musíme nadefinovat hodnotu  $d'$  způsobem, která nebude závislá na rozdělení veličiny  $X$ .

**Definice 2.** *Nechť má pozorovatel  $p$  hit rate  $H_p$  a false alarm rate  $F_p$  obě ostře mezi jedničkou a nulou. Potom budeme jeho senzitivitu značit  $d'$  a spočítáme jí jako  $d' = \Phi^{-1}(H_p) - \Phi^{-1}(F_p)$ , kde  $\Phi^{-1}$  je inverzní distribuční funkce normovaného normálního rozdělení.*



Obrázek 1.3: Závislost hit rate a false alarm rate pro různé hodnoty  $d'$ . Těmto křivkám se říká ROC křivky z anglického Receiver Operating Characteristic (Swets, 1996; Green a Swets, 1988).

Nahlédneme, že senzitivita má následující vlastnosti:

- Pokud  $d' = 0$ , tak jsou indikátorové náhodné veličiny „Pozorovatel odpověděl kladně“ a „Signál je přítomen“ nezávislé.
- Pokud  $d' > 0$ , tak mají tyto dvě veličiny kladnou korelaci.
- Pokud skutečně má náhodná veličina  $X$  jak v případě, že signál přítomen je, tak v případě, že signál přítomen není, normální rozdělení liší se pouze střední hodnotou, pak hodnota  $d'$  nezáleží na nastavení kritéria (samozřejmě krom případů, kdy kritérium nastavíme na plus nebo minus nekonečno, pak hodnota  $d'$  není dobře definovaná).

## 1.4 Entropie

Entropii (nazývanou též Shannonova entropie, aby nedošlo k záměně s entropií tak, jak jí chápe termodynamika) náhodné veličiny se v teorii informace rozumí střední hodnota informace, kterou nám hodnota této veličiny přinese. Například mějme náhodné veličiny  $X$  a  $Y$ .  $X$  nechť nabývá hodnoty 0 s pravděpodobností  $1/2$  a hodnoty 1 s toutéž pravděpodobností.  $Y$  nechť nabývá hodnoty 1 s pravděpodobností 1 a hodnoty 0 s pravděpodobností 0. Je vidět, že pokud se dozvíme, jaké hodnoty nabývá veličina  $X$ , získáme více informace, než když zjistíme, jaké hodnoty nabývá  $Y$ .

Hodnotu informace, kterou nám přineslo zjištění, že náhodná veličina  $A$  nabývá hodnoty  $a$ , spočítáme jako

$$I(A = a) = -\log_b (P[A = a]) = \log_b \left( \frac{1}{P[A = a]} \right).$$

Jako základ logaritmu  $b$  se běžně používá 2 (což budeme dělat i v této práci),  $e$  nebo 10. Všimneme si, že toto vyjádření množství informace je konzistentní s intuitivní představou, že zjištění, že  $A$  nabývá nějaké nepravděpodobné hodnoty, je cennější, než zjištění, že nabývá nějaké pravděpodobné. Odsud tedy entropii  $H(A)$  lze vyjádřit jako

$$H(A) = \mathbb{E}[I(A)] = - \sum_{a \in \Omega} P[A = a] \log_2 P[A = a],$$

kde  $\Omega = \{a | P[A = a] > 0\}$ . Všimneme si, že podle tohoto vzorce je entropie veličiny  $X$ , tak jak byla nadefinována v předchozím odstavci, rovna jedné, kdežto entropie veličiny  $Y$  je rovna nule.

Podobně jako u pravděpodobnosti můžeme poměrně přímočaře nadefinovat i podmíněnou entropii  $H(A|B)$

## 1.5 Modely pozorovatele

V této práci se budeme zabývat úlohou, kdy pozorovatel hledá gabor patch v kruhovém poli, v němž se nachází růžový šum. Předchozí výzkum ukázal, že lidští pozorovatelé v této úloze využívají krátkých pohledů a rychlých pohybů očí (Najemnik a Geisler, 2005). Jednomu takovému krátkému pohledu budeme říkat *fixace*. Očnímu pohybu se říká *sakáda* (Králíček, 2002). Pojem sakáda přirozeně

aplikujeme i na simulované (tj. ne liské) pozorovatele, kde ho budeme chápat jako vektor v  $\mathbb{R}^2$  mezi středy dvou po sobě jdoucích fixací. Jednotlivé fixace trvají přibližně 200–300 ms.

Abychom ale mohli hodnotit fixace lidského pozorovatele, potřebujeme mít nějaký model, který nám bude říkat, jak se chová optimální pozorovatel. Nejprve ale stručnou odbočku:

### 1.5.1 $d'$ mapa

Ještě předtím, než začneme zjišťovat, podle jakého modelu se chová lidský pozorovatel, je nutné zjistit, kolik informace člověk jednou fixací získá. Je například zjevné, že kdyby jedním pohledem bez ohledu na to, kam se dívá, získal stejné množství informace o všech možných polohách cíle, jsou všechny modely ekvivalentní. Proto je potřeba najít tzv.  $d'$  mapu, funkci, která nám pro každé dva body  $x, y$  řekne, jakou hodnotu  $d'$  má lidský pozorovatel, pokud hodnotí pozici  $y$  a dívá se na pozici  $x$ . U pozadí s uniformními nebo téměř uniformními lokálními kontrasty lze tuto funkci zjednodušit, stačí, když budeme pro každý bod  $x$  vědět, jaká je hodnota  $d'$  pro rozhodování, zda je cíl v bodě  $x$ , pokud zafixujeme střed.

Předchozí výzkum ukázal, že u běžného člověka je  $d'$  mapa poměrně nepravidelná (Najemnik a Geisler, 2008), ale dá se s přijatelnou přesností aproximovat mapou, kde křivky spojující body se stejnou hodnotou  $d'$  jsou tvořeny čtyřmi čtvrtelipsami (jednou v každém kvadrantu) se středem v počátku, s excentricitami a velikostí poloos danou hodnotou  $d'$  a individuálními vlastnostmi pozorovatele a cíle (Bradley a kol., 2014).

Ke kompletnímu vyjádření aproximace funkce  $d'$  potřebujeme celkem 6 hodnot, které závisí na konkrétním pozorovateli a cíli. Jedná se o hodnotu  $d'_0 = d'(0,0)$ , hodnot  $e_L$ ,  $e_R$ ,  $e_U$  a  $e_D$ , které určují vzdálenost ve čtyřech základních směrech takovou, že v nich je hodnota  $d'$  poloviční oproti počátku, a hodnotu  $\beta$  popisující sklon této funkce. Funkci  $d'$  pak vyjádříme jako

$$d'(x,y) = \frac{d'_0}{1 + \left( \frac{x^2}{e_H^2} + \frac{y^2}{e_V^2} \right)^\beta},$$

kde  $e_H$  je rovno  $e_L$  pro záporná  $x$  a  $e_R$  pro kladná, a  $e_V$  je rovno  $e_D$  pro záporná  $y$  a  $e_N$  pro kladná.

### 1.5.2 Modely chování pozorovatele

Všechny modely, které zde budeme zkoumat, vypadají tak, že mají tzv. *Mapu posteriorních pravděpodobností*. V této mapě je pro každou lokaci, kde by signál (též cíl) mohl být, uvedeno, jaká je pravděpodobnost, že se na ní cíl nachází vzhledem k informaci, kterou již o dané lokaci pozorovatel získal. Ve chvíli, kdy na nějaké lokaci pravděpodobnost přesáhne kritérium, pozorovatel ukončí hledání a ohlásí nalezení cíle na této lokaci.

V rámci předchozího výzkumu bylo otestováno mnoho modelů, jako například pozorovatel, který volí fixace náhodně nebo pozorovatel, který volí fixace tak, aby nikdy nezafixoval tutéž lokaci vícekrát. Všechny tyto modely se ale ukázaly jako nevhodné, vzhledem k tomu, že v praxi dosahují mnohem horších výsledků

(měřeno pomocí střední hodnoty počtu fixací před nalezením cíle) než lidský pozorovatel.

## MAP pozorovatel

Nejjednodušší model, který dosahuje podobných výsledků, jako lidský pozorovatel, je tzv. MAP<sup>4</sup> pozorovatel, který vždy zafixuje lokaci, která má v jeho mapě posteriorních pravděpodobností nejvyšší hodnotu. Tento pozorovatel již dosahuje podobných výsledků jako lidský, ale jeho strategie fixací neodpovídá strategii, jakou volí lidský pozorovatel. Lidský pozorovatel umístí svoji první fixaci do středu scény. Ostatní fixace jsou pak rozmístěny v okolí kružnice se středem ve středu scény a poloměrem rovným přibližně 2/3 poloměru scény, s preferencí pro horní a spodní okraj. MAP pozorovatel oproti tomu vybírá každou lokaci se zhruba stejnou pravděpodobností.

## Ideální Bayesovský pozorovatel

Ještě o malinko lepších a hlavně statisticky lidskému pozorovateli bližších výsledků obsahuje model Ideálního Bayesovského pozorovatele.

Ideální Bayesovský pozorovatel (dále IBO) je pozorovatel, který  $T + 1$  lokaci vybírá tak, aby maximalizoval pravděpodobnost, že v následujícím kroku odhalí cíl. Vybere tedy lokaci

$$k_{opt}(T + 1) = \arg \max_{k(T+1)} \left( \sum_{i \in L} p_T(i) P[p_{T+1}(i) \geq c | i, k(T + 1)] \right), \quad (1.1)$$

kde  $p_N$  je mapa posteriorních pravděpodobností po  $N$ -té fixaci,  $L$  je množina všech potenciálních lokací cíle a  $c$  je kritérium, které musí hodnota v mapě posteriorních pravděpodobností překročit, aby bylo ukončeno hledání a nahlášeno nalezení cíle. Výraz  $P[p_{T+1}(i) \geq c | i, k(T + 1)]$  pak tedy znamená „pravděpodobnost, že po  $T + 1$  fixaci bude ukončeno hledání a nhlášen signál v lokaci  $i$ , za podmínky, že tam signál opravdu je a byl zafixován bod  $k(T + 1)$ “.

IBO však ale též není příliš pravděpodobný kandidát na model, podle něž se lidé chovají. Ač má jeho vyhledávání podobné statistické vlastnosti, jako vyhledávání lidského pozorovatele, vyžaduje perfektní paměť a ideální integraci informace mezi fixacemi. Ani jednu z těchto dovedností ale lidský pozorovatel nemá. Dále si můžeme všimnout, že přinejmenším přímočaré vyhodnocení výrazu (1.1) je kubické v počtu potenciálních lokací (výpočet druhého činitele součinu v sumě je lineární, suma sama je přes lineárně mnoho členů, a vnější maximum má též lineárně mnoho možných voleb  $k(T + 1)$ ). Lidský pozorovatel při vyhledávání volí další fixaci přibližně třikrát až čtyřikrát za vteřinu, je tedy nepravděpodobné, že by lidský mozek dokázal takový výpočet provést (Najemnik a Geisler, 2008).

## ELM pozorovatel

ELM<sup>5</sup> pozorovatel je pozorovatel, který při výběru následující fixace minimalizuje střední hodnotu entropie náhodné veličiny určující lokaci cíle. Tu spočítáme

<sup>4</sup>Z anglického „Maximum A posteriori Probability“.

<sup>5</sup>Z anglického „Entropy limit minimization“.

jako

$$\mathbb{E} [H(T+1)|k(T+1)] = -\mathbb{E} \left[ \sum_{i \in L} p_{T+1}(i) \log_2 p_{T+1}(i) |k(T+1) \right]. \quad (1.2)$$

Najemnik a Geisler (2009) však ukázali, že vyjádření hodnoty entropie podle vzorce (1.2) je sice netriviální, ale lze ho dobře aproximovat případem, kdy pošleme v limitě  $|L|$  do nekonečna. Pak dostáváme výraz

$$\mathbb{E} [H(T+1)|k(T+1)] = H(T) - \sum_{i \in L} p_T(i) d'^2(i - k(T+1)),$$

kde v posledním členu bereme lokace jako vektory od počátku k nim. Člen  $H(T)$  navíc nezávisí  $k(T+1)$ , takže s ním vůbec nemusíme počítat a k minimalizaci entropie nám stačí maximalizovat hodnotu sumy.

## 2. Metody

Nemělo by být napřed formálně posáno, co to vlastně zkoumáme, než začneme popisovat metody?

### 2.1 Účastníci

### 2.2 Nástroje a stimuly

K experimentu byla použita aplikace, která je součástí této práce. Jako zobrazovací zařízení byl použit iPad air s displejem o rozlišení  $2048 \times 1536$  a o úhlopříčce 9.7 palců, což odpovídá rozměrům displeje asi  $19.7 \times 14.8$  centimetrů. Hustota pixelů je 264 pixelů na palec.

V experimentu použitý růžový šum byl kruhový a jeho průměr byl 1024 pixelů.

<sup>1</sup> Tento průměr byl zvolen proto, že je to nejbližší mocnina dvojky ke kratšímu rozměru displeje iPadu v pixelech. Mocnina dvojky byla zvolena kvůli použití rychlé Fourierova transformace při generování růžového šumu.

Jako stimulus byl použit Gabor patch. Nebyl ale přičten k šumu, ale vložen do něj, jako kdybychom kreslili Gabor patch přes šum a obálka zastupovala alfa kanál. Zvolené parametry Gabor patche byly:

- Obálka: Raised cosine
- Průměr: 50 pixelů
- Frekvence:  $1/16$  cyklu na pixel
- Fázový posun: 0
- Úhel  $\Theta$ :  $135^\circ$

Možných lokací Gabor patche bylo celkem 85, a byly rozmístěny po scéně v trojúhelníkové mřížce tak, aby jedna možná lokace byla ve středu. Vzdálenost dvou sousedních možných lokací byla 100 pixelů.

Kontrast cíle byl daný maximem obálky, tedy při snižování kontrastu byl cíl čím dál tím průhlednější. Hodnoty kontrastu se mohly pohybovat mezi nulou a jedničkou.

### 2.3 Procedura

Každý subjekt podstoupil napřed měření kontrastu, který musí gabor patch mít, aby pravděpodobnost, že bud detekován na středu zorného pole byla přibližně 90 % **Budeme to měřit na všech?**. Následně každý subjekt prošel sadou 3 testů. V prvním testu mu bylo postupně prezentováno 50 úkolů, kde v každém z nich měl najít Gabor patch v růžovém šumu. Úkol byl považován za úspěšný, pokud byl cíl nalezen během pěti či méně fixací. V druhém testu bylo prezentováno 150

---

<sup>1</sup>Odsud již všude, kde není zřejmý opak, se pixelem myslí pixel obrázku, nikoliv pixel displeje.



obdobných úkolů a ve třetím opět 50. Ve druhém testu dostávaly subjekty, které byly ve skupině se zpětnou vazbou, po každé fixaci zvukovou odpověď, která značila, kolik informace mohli od této fixace očekávat (tedy jestli bylo z pohledu IBO/ELM moudré udělat právě tuto fixaci). Tato odpověď byla ve formě tónu, jehož frekvence byla dána vzorcem

**Vymyslet,**

kde  $\Delta$  je maximální dosažitelné očekávané snížení entropie a  $\delta$  minimální v případě, že by byla zafixována některá z možných lokací cíle.<sup>2</sup> Pokaždé, když byl subjekt třikrát po sobě úspěšný, byla zvýšena obtížnost snížením kontrastu cíle **O kolik?**, pokud byl třikrát po sobě neúspěšný, byla obtížnost opět snížena.

V každém úkolu byl šum překryt černou barvou. Subjekt se měl vždy dotknout displeje v místě, které se rozhodl zafixovat. Na tomto místě byl poté šum odkryt na 300 ms. Výpočet tvaru a míry odkrytí oblasti bylo provedeno vynásobením s  $d'$  mapou posunutou do bodu fixace, s parametrem  $d'_0$  nastaveným na 1 a ostatními parametry naměřenými na pozorovateli FD ( $e_R = 223$ ,  $e_L = 223$ ,  $e_U = 161$ ,  $e_D = 164$ ,  $\beta = 2.46$ , všechny veličiny, u nichž má smysl uvádět jednotku, jsou v pixelech).

Ve chvíli, kdy si subjekt myslel, že objevil cíl, zmáčkl tlačítko. Poté mu byl ukázán celý odkrytý šum, ovšem bez cíle. Potom se měl subjekt dotknout šumu na místě, kde si myslel, že se cíl nacházel. Cíl byl považován za nalezený, pokud byla vzdálenost vybraného místa a středu skutečné lokace cíle menší než 50 pixelů.

## 2.4 Limitace

Při návrhu experimentu jsme narazili na několik problémů, které mohou vnést nepřesnosti do měření, ale jejichž řešení je mimo rozsah této práce. Konkrétně se jedná o následující obtíže:

- Vjemy lidského pozorovatele neodpovídají příliš dobře vjemům simulovaného ideálního pozorovatele. Aby si tyto vjemy odpovídaly, alespoň přibližně, museli bychom každému pozorovateli změřit jeho vlastní  $d'$  mapu. Měření  $d'$  mapy ale i v té nejminimalističtější variantě, která se používá, trvá nejméně jeden pracovní den. Druhým důvodem, proč mohou být vjemy rozdílné i v případě, že by konstanty  $d'$  mapy vyšly pozorovateli stejně, jaké byly použity, je, že tato naměřená  $d'$  mapa odpovídá situaci, kdy je scéna se šumem umístěna tak daleko od pozorovatele, aby ji viděl pod zorným úhlem  $15^\circ$ . To při velikosti scény v našem případě odpovídá vzdálenosti pozorovatele a zařízení přibližně 65 cm. V našich experimentech nebyla vzdálenost pozorovatele od scény hlídána, a určitě byla nižší, než řečených 65 cm (dodržení této vzdálenosti by odpovídalo situaci, kdy by subjekty držely iPad před sebou zhruba na délku natažené paže).

<sup>2</sup>To znamená, že je potenciálně možná dosáhnout výsledku lepšího než  $\Delta$ , resp. horšího, než  $\delta$ . Rozdíl by však neměl být důležitý. Meze jsou do vzorce přidány, abychom zabránili absurdním hodnotám v případě, kdy by si hodnoty  $\Delta$  a  $\delta$  byly téměř rovny, ale nějaký bod, který není možnou lokací, by měl očekávané snížení entropie o trochu vyšší či nižší.

- I pokud odhlédneme od nepřesností zmíněných v předchozím bodě a dovolíme si na chvíli (evidentně scestný) předpoklad, že subjekty měly vlastní  $d'$  mapu konstattní, narazíme na další problém. Okraj oblasti, která byla odkrývána, byl ztmavován lineárně se snižující se hodnotou  $d'$  v použité  $d'$  mapě. Závislost  $d'$  na kontrastu ale téměř jistě není lineární.
- S tím souvisí ještě jeden problém: Subjekty samozřejmě nemají svou  $d'$  mapu konstattní. Tato mapa se tedy nějak skládá s  $d'$  mapou, pomocí které bylo určeno odhalování šumu. V práci jsme toto skládání ignorovali (tedy předpokládali jsme, že  $d'$  na kontrastu závisí lineárně a  $d'$  mapa subjektů je konstattní.) Nabízela by se otázka, proč tedy bylo zatemňování šumu vůbec prováděno. To se dělo z několika důvodů:
  - Zatemňování šumu zavádí potřebu klikat na místa, která chce pozorovat v dalším kroku prozkoumat. Nutí tedy pozorovatele, aby tato rozhodnutí dělal vědomě a nikoli podvědomě, což byl jeden z efektů, které jsme chtěli zkoumat.
  - Celý proces jedné fixace tímto způsobem také trvá mnohem déle (nižší jednotky vteřin místo nižších desetin vteřiny) a poskytuje nám tedy mnoho času na update mapy posteriorních pravděpodobností a výpočet množství informace, kterou lze získat následující fixací.
  - Takto navržený experiment též umožňuje zjišťovat, které lokace subjekt fixuje bez použití eyetrackeru nebo jiných technologií.
- Je možné, že u lidí, kteří nikdy dříve Gabor patch hledat nezkoušeli, se během testů mění jejich sensitivita na Gabor patch (mění se hodnota  $d'(0,0)$ ).
- Vzhledem k tomu, že počet reálných pixelů displeje neodpovídal (a ani nebyl dělitelný) velikostí scény v pixelech, je možné, že byly efekty jako například antialiasing změněny lokální kontrasty scény.
- Ve výzkumu v oblasti psychofyziky se většinou pečlivě kontroluje prostředí (například se zatemňuje místnost, v níž se provádí experiment). To jsme v našem výzkumu nedělali.

Tyto nepřesnosti však nepovažujeme za příliš důležité, protože cílem této práce není přesný experiment, ale pouze proof of concept. Za zásadní chybu naopak nepovažujeme malý počet účastníků – kdybychom chtěli na této práci postavit přesný experiment, nebylo by potřeba zvyšovat počet účastníků, ale pouze počet měření na jednotlivých účastnících, například tak, že bychom vícekrát opakovali první a třetí test. Jinak ale není ve výzkumu nutně přínosné velký vzorek, často je lepší na malém vzorku provést větší počet přesnějších měření (Smith a Little, 2018).

## **3. Měření**

### **3.1 Metodika**

### **3.2 Výsledky**

## 4. Implementace

Tady bych dal nejspíše nějaký high-level přehled a potom pár slov o každém zdrojovém souboru.

# Závěr

# Seznam použité literatury

- AMEMORI, J. (2016). Ideální bayesovský pozorovatel s redukovanou detekční mapou. Diplomová práce. Univerzita Karlova, Matematicko-fyzikální fakulta, Katedra softwaru a výuky informatiky. Vedoucí práce Děchtěrenko, Filip.
- BRADLEY, C., ABRAMS, J. a GEISLER, W. S. (2014). Retina-V1 model of detectability across the visual field. *Journal of Vision*, **14**(12), 22. doi: 10.1167/14.12.22. URL <http://dx.doi.org/10.1167/14.12.22>.
- FOGEL, I. a SAGI, D. (1989). Gabor filters as texture discriminator. *Biological Cybernetics*, **61**(2), 103–113. ISSN 1432-0770. doi: 10.1007/BF00204594. URL <https://doi.org/10.1007/BF00204594>.
- GREEN, D. a SWETS, J. (1988). *Signal Detection Theory and Psychophysics*. Peninsula Pub. ISBN 9780932146236. URL <https://books.google.cz/books?id=AjWwQgAACAAJ>.
- HEEGER, D. (1997). Signal detection theory. URL <http://www.cns.nyu.edu/~david/handouts/sdt-advanced.pdf>.
- KRÁLÍČEK, P. (2002). *Úvod do speciální neurofyzologie*. Karolinum. ISBN 80-246-0350-0.
- NAJEMNIK, J. a GEISLER, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, **434**, 387–391.
- NAJEMNIK, J. a GEISLER, W. S. (2008). Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision*, **8**(3), 4. doi: 10.1167/8.3.4. URL <http://dx.doi.org/10.1167/8.3.4>.
- NAJEMNIK, J. a GEISLER, W. S. (2009). Simple summation rule for optimal fixation selection in visual search. *Vision Research*, **49**(10), 1286 – 1294. ISSN 0042-6989. doi: <https://doi.org/10.1016/j.visres.2008.12.005>. URL <http://www.sciencedirect.com/science/article/pii/S0042698908006007>. Visual Attention: Psychophysics, electrophysiology and neuroimaging.
- SMITH, P. L. a LITTLE, D. R. (2018). Small is beautiful: In defense of the small-n design. *Psychonomic Bulletin & Review*. ISSN 1531-5320. doi: 10.3758/s13423-018-1451-8. URL <https://doi.org/10.3758/s13423-018-1451-8>.
- SWETS, J. (1996). *Signal detection theory and ROC analysis in psychology and diagnostics: collected papers*. Scientific psychology series. Lawrence Erlbaum Associates. ISBN 9780805818345. URL <https://books.google.cz/books?id=1vkMAQAAMAAJ>.

## Problémy

19: Tyhle obecný kecy moc nikde na jednom místě krom Wikipedie nejsou – něco málo je tady <http://www.atis.org/glossary/> – Nějaká americká telekomunikační norma.

19: Nemělo by být napřed formálně posáno, co to vlastně zkoumáme, než začneme popisovat metody?

19: Budeme to měřit na všech?

19: O kolik?

# Seznam obrázků

1.1	Ukázky různých šumů. . . . .	4
1.2	Ukázky několika Gabor patchů. Všechny gabor patche jsou 100 pixelů široké i vysoké. Levý patch má $\Theta = 1/4\pi$ , ostatní mají $\Theta = -1/4\pi$ , levý má jako obálku Gaussovu funkci, prostřední dva raised cosine, pravý lineární funkci vzdálenosti, první, druhý a čtvrtý mají frekvenci (v cyklech na pixel) 0.1, třetí 0.02. . . . .	5
1.3	Závislost hit rate a false alarm rate pro různé hodnoty $d'$ . Těmto křivkám se říká ROC křivky z anglického Receiver Operating Characteristic (Swets, 1996; Green a Swets, 1988). . . . .	7



# Seznam použitých zkratek

## A. Přílohy

### A.1 První příloha