

卒業論文

ビデオ映像からの  
文字列抽出手法に関する研究

東北大学 工学部 通信工学科  
阿曾研究室 4年  
野田口 玄

# 目次

<b>第1章</b>	<b>序論</b>	<b>1</b>
1.1	研究の背景	1
1.2	研究の目的	2
1.3	本論文の構成	3
<b>第2章</b>	<b>松尾らの手法</b>	<b>4</b>
2.1	概要	4
2.2	分解画像の作成	5
2.2.1	HSL 変換	5
2.2.2	情景画像の2値化	5
2.2.3	等間隔分解画像	6
2.3	文字領域候補の抽出	7
2.3.1	ラベリングによる外接矩形抽出	7
2.3.2	外接矩形からの文字抽出条件	7
2.3.3	マーキング処理	7
2.4	文字列候補の抽出	9
2.4.1	文字列候補領域の判断	9
2.4.2	文字列領域の抽出	10
<b>第3章</b>	<b>提案手法</b>	<b>11</b>
3.1	提案手法	11
3.2	動画画像から画像を1枚切り出して文字列抽出を行う手法	11
3.2.1	メディアンフィルタ	12
3.2.2	膨張・縮小	12
3.3	動画画像から画像を3枚切り出して文字列抽出を行う手法	13
3.3.1	2値画像の重ね合わせ	13
<b>第4章</b>	<b>実験</b>	<b>16</b>
4.1	はじめに	16
4.2	実験条件	16
4.3	実験	17
4.4	実験結果の考察	18
4.4.1	1枚の切り出し画像から文字列抽出を行う手法	18

4.4.2	3枚の切り出し画像から文字列抽出を行う手法 . . . . .	21
<b>第5章</b>	<b>結論</b>	<b>22</b>
5.1	まとめ . . . . .	22
5.2	今後の課題 . . . . .	22
	<b>謝辞</b>	<b>23</b>
	<b>参考文献</b>	<b>24</b>

## 目次

2.1	松尾らの手法	4
2.2	明度ヒストグラム	6
2.3	彩度色相分布図	6
2.4	左：交差矩形マーキング処理　右：分離矩形マーキング処理	7
2.5	条件1	9
2.6	条件2	9
2.7	条件3	10
3.1	切り出し画像を1枚用いた手法	12
3.2	左：膨張処理　右：縮小処理	12
3.3	連続した切り出し画像を3枚用いた手法	13
3.4	連続した3枚の画像	14
3.5	連続した3枚の画像のある明度ヒストグラムの範囲での2値画像	14
3.6	それぞれの看板の領域の拡大画像(2倍)	15
3.7	連続した3枚の画像を重ね合わせた画像	15
4.1	ノイズの影響による色の侵食	18
4.2	図(4.1)の部分的な拡大(3倍)	18
4.3	情景画像	19
4.4	膨張→縮小未使用	19
4.5	膨張→縮小使用時の失敗例	20
4.6	縮小→膨張使用時の失敗例	20

# 第1章

## 序論

### 1.1 研究の背景

現在、パターン認識の分野では様々な研究が行われ、視覚を用いた認識を行う移動ロボットの実現が可能となってきた。現段階の研究では、障害物などの比較的単純な形をした物体の認識は比較的容易に行うことができる。しかし、文字や記号などの重要な情報を保持しているものは、情景中の細かな物体などと混ざってしまい、認識は非常に困難なものである。

しかし、特に文字は情報の伝達性や保存性に優れており、あらゆる情報を様々な場所で提供する重要なものである。障害物だけでなく、文字をきちんと認識することができれば、より高度な移動ロボットの実現が可能となるはずである。

人間であれば、文字領域を用意に判断し、即座に認識・理解することが可能であるが、移動ロボットが文字を自動的に認識する場合には、文字の方向、大きさ、形状、また画像の明るさや解像度などが問題となる。

従来行われてきた情景画像からの文字列領域の抽出の手法は、恩田らのドアの部屋番号の文字列領域抽出 [1] や、藪木らの道路情景画像からの速度標識の検出 [2]、松尾らの色彩情報を用いた情景画像からの文字列抽出 [3] などがあるが、恩田ら、藪木らの手法では背景領域もしくは文字領域が均一色であったり、文字領域の書式が知識として与えられているという特徴があり、また松尾らの手法では、情景画像そのものから文字列の抽出をしてはいるが、デジタルカメラからの静止画像を用いているという特徴がある。

移動ロボットというものは様々な情報をリアルタイムで取り入れるため、どの手法も少々不向きな点があると言えるが、情景画像そのものから文字列抽出を行うという点では松尾らの手法が最も近い手法と言える。

ここで言うリアルタイムというのは、認識の点でのリアルタイムではなく、情報を取り入れる点でのリアルタイムということである。

本論文では、移動ロボットは複雑な情景中の動画像の情報をリアルタイムで取り入れ認識を行うので、松尾らの手法を参考にし、デジタルカメラによる静止画像からの文字列抽出とは異なる、抽出の際の問題点を考慮したビデオ映像からの文字列抽出手法を提案する。

今回の手法では、松尾らの手法と同様にビデオ映像から画像を1枚切り出し、静止画像と同じように扱い文字列抽出を行い、さらにビデオ映像を用いることによるノイズの影響も考慮した手法と、ビデオ映像であることを考慮し、さらに前後1枚ずつの画像、つまり合計3枚の連続した画像を用いた手法の提案をする。

## 1.2 研究の目的

本研究では、ビデオ映像を用いることによるノイズの影響を考慮した情景画像からの文字列抽出を行うことを目的としている。

## 1.3 本論文の構成

本論文の構成は以下の通りである。

### 第1章 序論

研究の背景や目的, 論文の構成を述べる。

### 第2章 松尾らの手法

今回参考にする松尾らの手法について述べる。

### 第3章 提案手法

本研究におけるビデオ映像からの文字列抽出に関する手法を述べる。

### 第4章 実験

抽出実験の結果と考察について述べる。

### 第5章 結論

本研究の成果, 今後の課題について述べる。

## 第2章

# 松尾らの手法

### 2.1 概要

まず、松尾らの手法について詳しく述べる。松尾らの手法は図(2.1)のように大きく分けて3段階に分けられる。まず初めに分解画像の作成を行い、次に文字領域候補の抽出、最後に文字列領域候補の抽出を行う。分解画像の作成の段階では、HSL変換を行い、2値画像の作成を行う。文字列領域候補の抽出の段階ではラベリング、外接矩形枠作成、マーキングを行う。文字列領域候補の抽出の段階では文字列領域の抽出を行う。

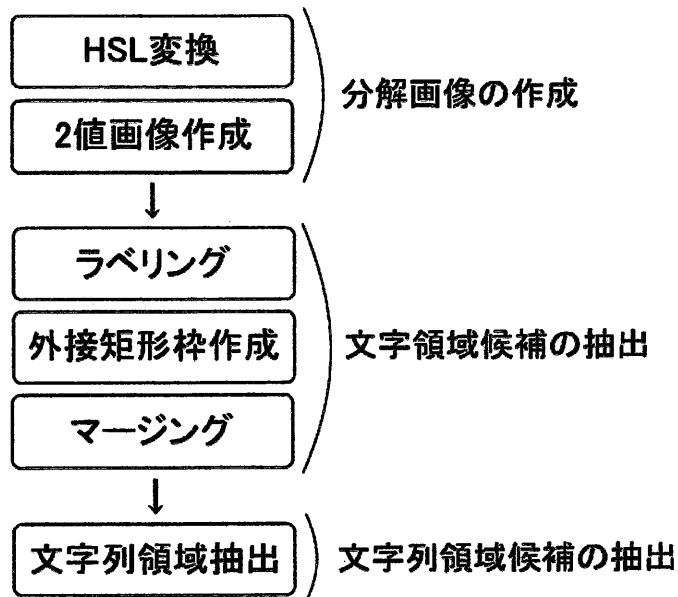


図 2.1: 松尾らの手法

次に松尾らの手法について各段階ごとにさらに詳しく述べる。



## 2.2 分解画像の作成

### 2.2.1 HSL 変換

情景画像をRGB信号に分解して入力した後に、HSL変換により、R、G、Bの輝度信号を表す明度 (Lightness)、色の違いを表す色相 (Hue)、そして色の鮮やかさを現す彩度 (Saturation) の3つの属性に変換する。

つまり、式(2.1)によって、RGBの輝度信号から明るさを表す明度値Lを求める。次に、色の鮮やかさを表す尺度として、式(2.2)より彩度値Sを求める。彩度値Sは、値が大きいと色味が増し、逆に小さくなると色味が失われる。このとき、彩度値Sが0で色味のない無彩色となる。色相は、式(2.3)のように、角度によって色の違いを表現する。

$$L = \frac{(R + G + B)}{3} \quad (2.1)$$

$$S = \frac{\text{Max}(R, G, B) - \text{Min}(R, G, B)}{\text{Max}(R, G, B)} \times 255 \quad (2.2)$$

Max(R,G,B)=Rの時

$$H = \frac{G - B}{\text{Max}(R, G, B) - \text{Min}(R, G, B)} \times 60$$

Max(R,G,B)=Gの時

$$H = \left( \frac{B - R}{\text{Max}(R, G, B) - \text{Min}(R, G, B)} + 2 \right) \times 60 \quad (2.3)$$

Max(R,G,B)=Bの時

$$H = \left( \frac{R - G}{\text{Max}(R, G, B) - \text{Min}(R, G, B)} + 4 \right) \times 60$$

### 2.2.2 情景画像の2値化

HSL変換により、情景画像を明度、彩度、色相の3つの特徴に分類する。これらの各特徴を、画像内の各画素について求め、横軸に明度値L、縦軸に横軸の明度値Lに対応した明度値の累積値を取ったものを図(2.2)に示すような明度ヒストグラムとする。また、色相値Hを横軸に、彩度値Sを縦軸の尺度にとったものを図(2.3)に示すような彩度色相分布図とする。

この明度ヒストグラムや彩度色相分布図そのものから文字領域がどこに分布しているかを判断するのは困難であるが、文字のもつ明度および色彩情報は、明度ヒストグラムや彩度色相分布図のある場所に必ず存在するので、明度ヒストグラムと彩度色相分布図を決められた領域ごとに分割し、分割された領域において得られる画像の閉領域の中に、対象となる文字列領域が存在すると考えられる。

### 2.2.3 等間隔分解画像

明度ヒストグラムを等間隔に8分割する。明度値Lは256階調であるから、分割された1領域の明度範囲は32となる。これにより分解された領域の範囲の明度値が閉領域となる2値画像を8枚得る。

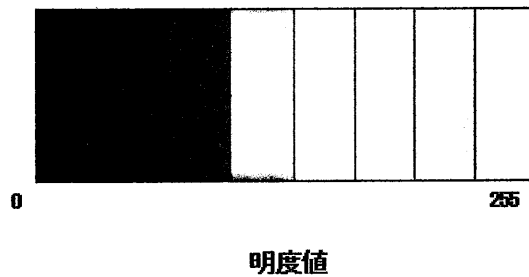


図 2.2: 明度ヒストグラム

次に、色彩情報に対しては、彩度色相分布図上で図1に示すように彩度値30未満を無彩色領域とし、彩度値30以上では、色相値0度から360度までを60度刻みで分割し、2値画像を計7枚得る。

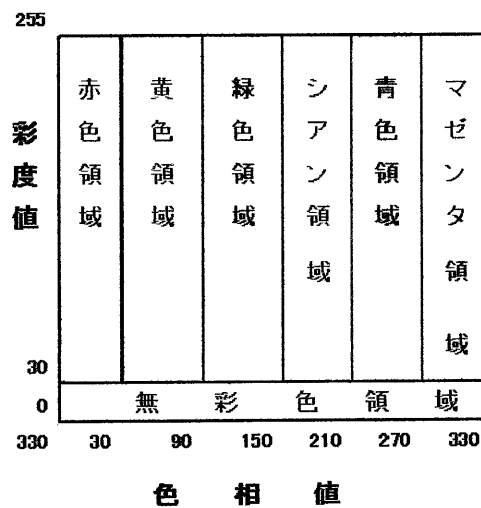


図 2.3: 彩度色相分布図

## 2.3 文字領域候補の抽出

### 2.3.1 ラベリングによる外接矩形抽出

文字や背景などの対象領域を抽出するために、ここではラベル付けされた閉領域の外接矩形を用いる。つまり、明度値と彩度・色相値によって得られた各分解画像に対して、同一の画素値をもつ閉領域にラベルを8連結で与えていき同じ領域として統合する。これによりラベルによる分解画像の領域分割が行える。この処理と並行して、ラベルによって分割された閉領域の  $x$ ,  $y$  座標の各最大値と最小値を求める。この各座標をもとに閉領域の外接矩形が得られる。

### 2.3.2 外接矩形からの文字抽出条件

ラベル付け処理によって得られた閉領域が対象となる文字領域であるか、もしくは文字を構成している一部分であるかを外接矩形の形状特徴により判断する。ここでは文字領域候補と文字部分領域候補の抽出条件を次のように定義する。

- a. 外接矩形の一辺が画素数  $p$  以上かつ縦横の辺比率が 1.5 未満の矩形は「文字領域候補」。
- b.aにおいて、一辺が画素数  $p$  以上かつ縦横の辺比率が 1.5 以上 10 以下の矩形は「文字部分領域候補」。
- c.a または b 以外の矩形は「不用」とし、対象としない。

上記の条件により抽出された外接矩形は「文字領域候補」、「文字部分領域候補」、「不要」に分類される。また試行実験により画素数を  $p=5$  とする。

### 2.3.3 マージング処理

マージングとは、図(2.4)で示すようなある二つの領域の外接矩形の形状と交差状態に基づき、その領域を統合、包括する処理である。ここでは、矩形の交差状態により2段階のマージングを行う。一つは重なった矩形同士のみをマージングする交差矩形マージング処理。もう一つは分離した矩形同士をマージングする分離矩形マージング処理である。交差矩形マージング処理では、矩形同士が交差または包括状態である場合にその両矩形を囲む新たな矩形を作成する処理である。分離矩形マージング処理は、偏と旁矩形が分離して存在する文字領域に対してこれを一つの矩形に統合する処理である。



図 2.4: 左 : 交差矩形マージング処理      右 : 分離矩形マージング処理

しかし、無条件に交差・分離マージング処理を続けると、文字領域候補の矩形が不要の矩形になったり、矩形が立て続けにマージされるなどの問題が起こる。そこでこれに対しマージング後の矩形の状態と、現在の矩形の状態によりマージング処理の実行を判断する条件を付加する。具体的には、マージング前の二つの外接矩形を $\alpha$ 、 $\beta$ として矩形 $\alpha$   $\beta$ を辺比率により、文字を構成している3種類の矩形（文字領域候補、文字部分領域候補、不要）に分類する。このとき、3種類の矩形の分類状態により、表1に示すようにどの候補同士が交差しているかを分類しランク分けを行う。ランクはA, B, Cの3種類とする。

矩形 $\alpha$ 矩形 $\beta$	文字領域 候補	文字部分 領域候補	不用
文字領域候補	ランクA	ランクB	ランクC
文字部分領域候補	ランクB	ランクA	ランクC
不用	ランクC	ランクC	ランクC

表 2.1: 矩形形状によるランク分け

次に、矩形 $\alpha$ と $\beta$ のマージング後の矩形状態に着目し、表2に示すランク分けされた3状態とマージング処理後の結果から、矩形 $\alpha$ と矩形 $\beta$ のマージング処理の実行を決定する。

結合後の矩形状態 矩形のランク	文字領域 候補	文字部分 領域候補	不用
ランクA	可	可	不可
ランクB	可	不可	不可
ランクC	不可	不可	不可

表 2.2: ランクによるマージングの決定

これによって、マージングの実行が決まると、矩形 $\alpha$ のラベル上のある画素のRGB値を $L_\alpha(R)$ ,  $L_\alpha(G)$ ,  $L_\alpha(B)$ とし、ラベルの面積を $P_\alpha$ として、式(2.4)よりラベル内の平均色 $\overline{L}_\alpha(R)$ ,  $\overline{L}_\alpha(G)$ ,  $\overline{L}_\alpha(B)$ を求める。そして、先に述べたマージング処理で、矩形 $\alpha$ と矩形 $\beta$ 内のラベルの平均色の色差 $D(\alpha, \beta)$ を式(2.5)で表すとき、この色差 $D(\alpha, \beta) \leq e$ のものについてマージング処理を実行する。

$$\begin{aligned}\overline{L}_\alpha(R) &= \frac{\sum L_\alpha(R)}{P_\alpha} \\ \overline{L}_\alpha(G) &= \frac{\sum L_\alpha(G)}{P_\alpha} \\ \overline{L}_\alpha(B) &= \frac{\sum L_\alpha(B)}{P_\alpha}\end{aligned}\tag{2.4}$$

$$D(\alpha, \beta) = |\overline{L}_\alpha(R) - \overline{L}_\beta(R)| + |\overline{L}_\alpha(G) - \overline{L}_\beta(G)| + |\overline{L}_\alpha(B) - \overline{L}_\beta(B)| \quad (2.5)$$

色差の有無は彩度色相分布図上の2つの領域にまたがらず、かつ人間が識別が困難である範囲として、今回の実験では $e=30$ と設定する。これにより文字領域候補と他の領域矩形との併合を防止することができる。

## 2.4 文字列候補の抽出

### 2.4.1 文字列候補領域の判断

情景画像中の文字部分に着目すると、文字が単独で存在することは稀である。むしろ文字は文になっていたり文字列として存在する。このことから、文字と文字との連結性を文字列抽出条件とすることができる。矩形レベルでは文字でない雑音も文字領域候補の矩形となるため、図(2.5)から図(2.6)に示す条件をもとに文字の連結性を調べ、文字列領域候補を抽出する。各分解画像で得られた全ての文字領域候補の矩形に対し、4つの文字列の条件が成り立つものを調べ、4つの条件がすべて成立するものを文字列を構成している矩形とみなし、これらの矩形を文字列領域候補と定義して抽出する。

1. 文字領域候補の矩形同士の面積比率が1.5以下である。

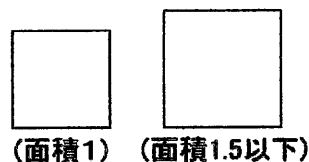


図 2.5: 条件 1

2. 一つの文字領域候補の矩形が水平または垂直移動したと仮定し、その移動の軌跡上にもう一つの矩形の重心が存在する。

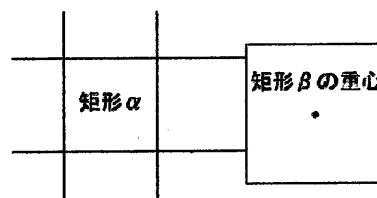


図 2.6: 条件 2

3. 文字領域候補の矩形間の距離が矩形3文字分以下である。

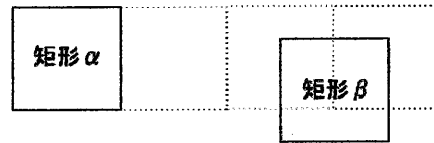


図 2.7: 条件 3

4. 文字領域候補の矩形内の対象領域における RGB 平均値の色差  $D(\alpha, \beta) \leq e$  である。

### 2.4.2 文字列領域の抽出

各分解画像が得られたら文字列領域候補から文字列領域を決定する。まず分解画像から抽出された全ての文字列領域候補を文字列統合画像に書き込む処理を行う。これにより、文字列統合画像は各分解画像から抽出された文字列領域候補を全て重ね合わせた画像となる。この画像から縦、または横に存在する文字列領域を抽出するために、文字列領域候補の個数を1つとみなしX軸、またはY軸に投影する。このとき、X、Y軸には文字列領域候補が交わった総数値を表すヒストグラムが得られる。この文字総数ヒストグラムは、分解画像上で同じ領域に文字列領域候補が得られるほど、文字総数ヒストグラム値が高くなる。つまり文字列領域の抽出には、この文字総数ヒストグラムを用いて以下の手法により行う。まず、画像サイズの一辺が  $m$  のときのX、Y座標の文字総数ヒストグラムを  $H_x$  とし、文字総数ヒストグラム値が0である総数を  $H_0$  とするとき、平均文字数  $M_c$  を式(2.6)によって求める。

$$M_c = \frac{\sum_{x=0}^{m-1} H_x}{m - H_0} \quad (2.6)$$

そして式(2.7)により文字総数ヒストグラム値  $H_x$  から、平均文字数  $M_c$  を差し引いた新たな文字総数ヒストグラム  $H'_x$  を求める。

$$H'_x = \begin{cases} H_x - M_c & H_x \geq M_c \\ 0 & H_x < M_c \end{cases} \quad (2.7)$$

この新たな文字総数ヒストグラム  $H'_x$  に存在するすべてのヒストグラムの山に対し、山の頂点から谷に向かって下るとき、文字総数ヒストグラムのX、Y座標上の極小値  $H_{xmin}$  と極大値  $H_{xmax}$  に囲まれたX、Y座標の範囲内に、存在する文字列領域候補をすべて囲みうる領域が最終的に文字列領域となる。

## 第3章

# 提案手法

### 3.1 提案手法

今回の提案手法では、松尾らの手法を参考にしながら、それを元に新たな手法を提案し、それに加えて文字列領域抽出の実験を行う。

### 3.2 動画像から画像を1枚切り出して文字列抽出を行う手法

まず、動画像から画像を1枚切り出して文字列抽出を行う手法では、図(3.1)のように第一段階の分解画像の作成の段階で新たな手法を用いる。一つはメディアンフィルタという主に画像の平滑化に用いる処理である。もう一つは、膨張・縮小という2値画像の領域補正に用いられる処理である。どちらの処理も、画像処理の方法としては一般に広く知られている処理である。この二つの手法について詳しく述べる。

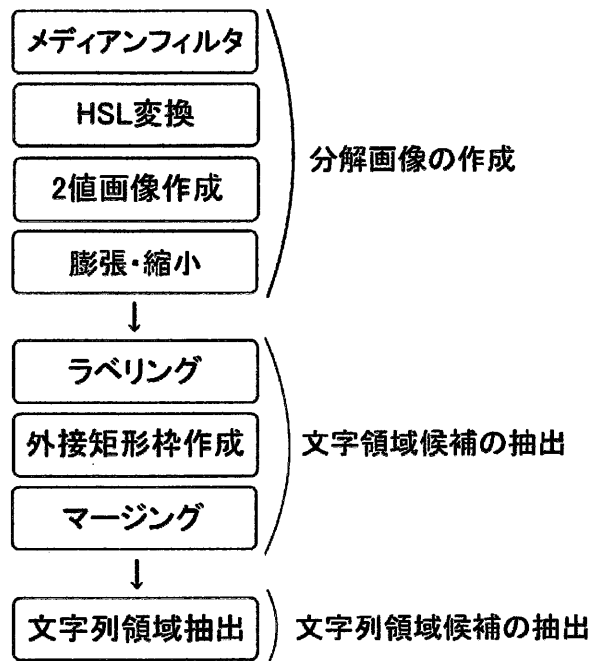


図 3.1: 切り出し画像を1枚用いた手法

### 3.2.1 メディアンフィルタ

メディアンフィルタとは、中心の画素とその周囲8画素の濃度の中央値を中心の画素の値にする処理である。この処理は画像のノイズとなる画素を取り除く処理で、画像の平滑化に効果がある処理である。この処理を行うことで、ビデオ映像により生じるノイズの影響を抑えることができるのではないかと考えられる。今回は試行実験によりフィルタを5回使用するものとする。

### 3.2.2 膨張・縮小

まず膨張について述べると、膨張とは有色画素を周囲8近傍に膨張させる処理で、2値画像において途切れた線などを繋ぐ効果がある。縮小は、逆に有色画素を1画素分内側に縮小させる処理で、2値画像において白黒の微小面積の領域を除去する。

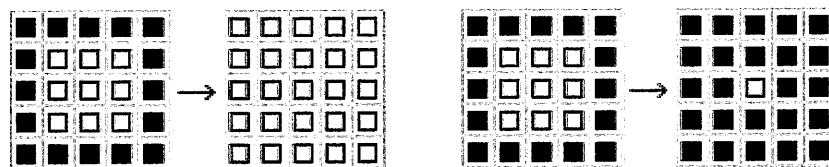


図 3.2: 左：膨張処理 右：縮小処理



### 3.3 動画画像から画像を3枚切り出して文字列抽出を行う手法

次に、動画画像から画像を3枚切り出して文字列抽出を行う手法では、図(3.3)のように第一段階の処理を三枚の画像に対して行う。ここで述べている三枚の画像というのは、動画画像のある1フレームの1画像の、前後のフレーム1枚ずつのことである。ここで三枚の画像を用いるのは、ビデオ映像から生じるノイズの影響で画像を切り出したときにもノイズが生じてしまうが、1フレームごとにノイズは異なるので、三枚の画像の2値画像を重ね合わせてできるだけノイズの影響の少ない2値画像にするのが目的である。

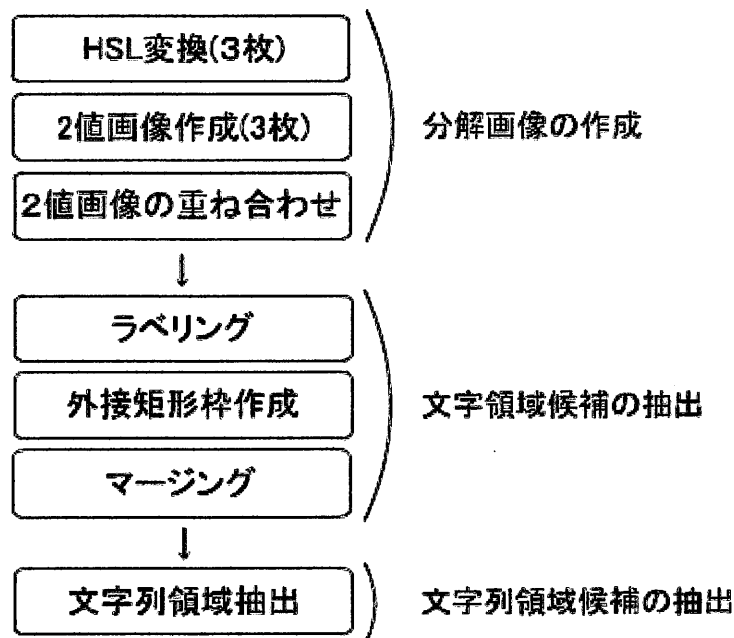


図 3.3: 連続した切り出し画像を3枚用いた手法

#### 3.3.1 2値画像の重ね合わせ

実際の処理は図(3.4)のような処理となる。まず、連続した3フレームの画像それぞれについて個別にHSL変換を行い、2値画像をそれぞれ作成し、最後にその画像を各範囲ごとに重ね合わせる。重ね合わせの条件は、三枚の画像の同じピクセル位置で領域が重なっていたならばその領域を重ね合わせて新たな重ね合わせた画像を作成する。

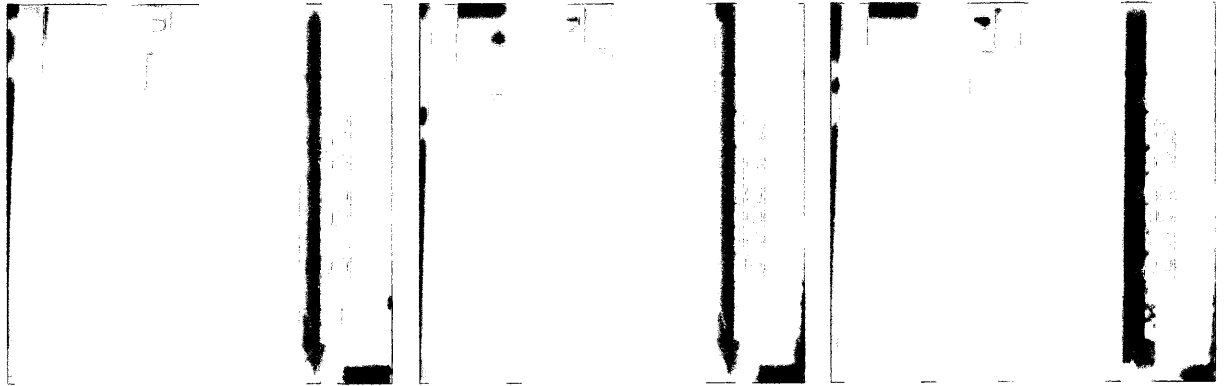


図 3.4: 連続した3枚の画像

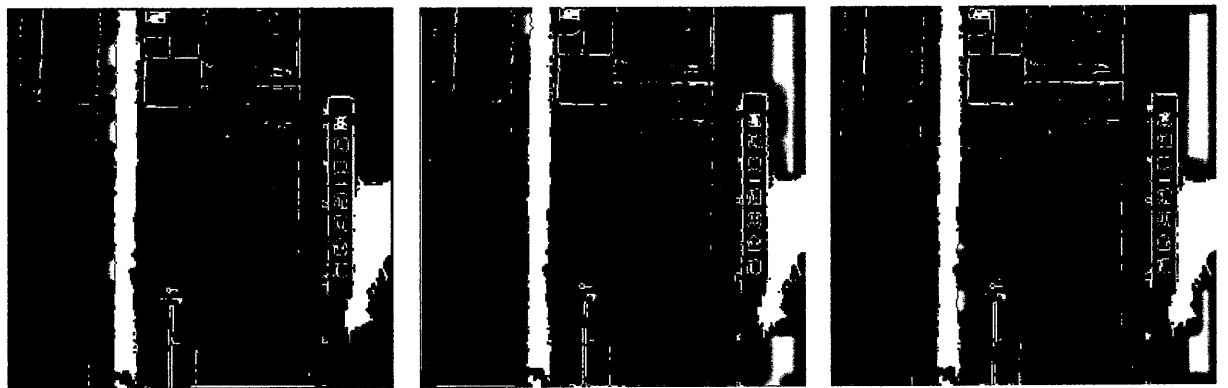


図 3.5: 連続した3枚の画像のある明度ヒストグラムの範囲での2値画像

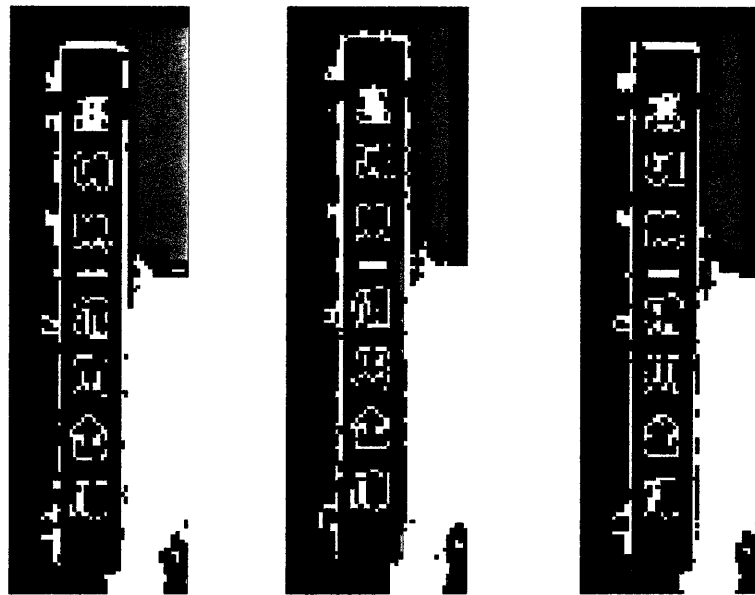


図 3.6: それぞれの看板の領域の拡大画像 (2倍)

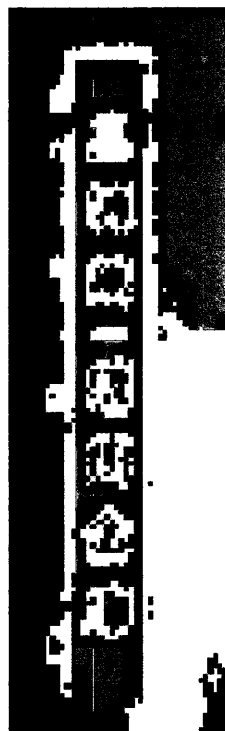


図 3.7: 連続した3枚の画像を重ね合わせた画像

## 第4章

# 実験

### 4.1 はじめに

今回の実験の詳細を述べる。

### 4.2 実験条件

実験では Canon デジタルビデオカメラ XV2 を用いて撮影を行い、縦 256 × 横 256 画素、RGB 各 256 階調の 20 枚のカラー情景画像データについて実験を行った。撮影条件は、比較的天気の良い日中に行い、対象文字列には飾り文字や手書き文字が含まれていないものとする。実験は

- 動画像から画像を 1 枚切り出して文字列抽出を行う手法
  - － 松尾らの手法
  - － 膨張→縮小処理を組み合わせたもの
  - － 膨張→縮小，縮小→膨張処理を組み合わせたもの
  - － メディアンフィルタを組み合わせたもの
  - － メディアンフィルタ，膨張→縮小処理を組み合わせたもの
  - － メディアンフィルタ，膨張→縮小，縮小→膨張処理を組み合わせたもの
- 動画像から画像を 3 枚切り出して文字列抽出を行う手法

の 7 つの条件で行う。

また、文字列抽出の判断は目視により行うが、その際に抽出結果にランクを付けて評価を行う。

- A : 文字列領域を全て抽出
- B : 文字列領域を他領域も含めて抽出
- C : 文字列領域を部分的に抽出
- D : 文字列領域を他領域も含めて部分的に抽出
- E : 他領域を抽出、または何も抽出できず

上記の5つにランク分けを行い5段階の評価を行う。

### 4.3 実験

二つの手法で実験を行った結果を以下に示す。

手法 抽出率	松尾	膨→縮	膨→縮 縮→膨	メ	メ、膨→縮	メ、膨→縮 縮→膨	3枚画像
A	30%	25%	25%	30%	35%	25%	45%
B	15%	20%	10%	15%	15%	10%	20%
C	25%	35%	35%	35%	20%	20%	15%
D	25%	5%	20%	20%	10%	25%	10%
E	5%	15%	10%	0%	20%	20%	10%
合計	100%	100%	100%	100%	100%	100%	100%

表 4.1: 文字列領域の抽出率

表(4.1)を見ると、松尾らの手法に対して、1枚の画像に処理を加えた手法では、あまり変化がないか、もしくは抽出率が悪くなっているのが分かる。それに対して3枚の連続した画像の2値画像を重ね合わせた手法では、ランクが高く抽出率も高いのが分かる。

次に、20枚の画像の1枚1枚に対して、松尾らの手法と比較して提案手法のランクがどう変化したかを表(4.2)として示す。

手法 ランクの変化	膨→縮	膨→縮 縮→膨	メ	メ、膨→縮	メ、膨→縮 縮→膨	3枚画像
向上	25%	10%	25%	25%	10%	25%
変化無し	40%	45%	50%	35%	45%	55%
低下	35%	45%	25%	40%	45%	20%
合計	100%	100%	100%	100%	100%	100%

表 4.2: 松尾らの手法と提案手法の比較

表(4.2)を見ると、全体的に似通った結果になっているように見えるが、ランクの向上が低下を上回っているものが3枚の画像を用いているものだけだということが分かる。1枚の画像に処理

を加えた手法は松尾らの手法に比べて全て抽出率が低下していることが分かる。ただし、ランクが同じ場合でも、抽出した文字数が多い場合には向上、抽出した文字数が少ない場合には低下としている。

## 4.4 実験結果の考察

### 4.4.1 1枚の切り出し画像から文字列抽出を行う手法

実験を行った結果、抽出率は全体的に変化が無いが低下であった。この原因としては、以下のことが挙げられる。

ノイズによる影響で、メディアンフィルタによる平滑化は逆効果であることが分かった。メディアンフィルタというものは、画像の平滑化をする際に、少々のノイズを取り除く処理に効果があり、図(4.1)の拡大図である図(4.2)のようにノイズの影響で看板の背景の色が文字の部分にまで侵食しているような画像では、逆に背景の色が混ざっている部分が多い領域では背景の色に平滑化されてしまい、結果として文字の領域がさらに少なくなってしまい領域として抽出できなくなるものもあった。

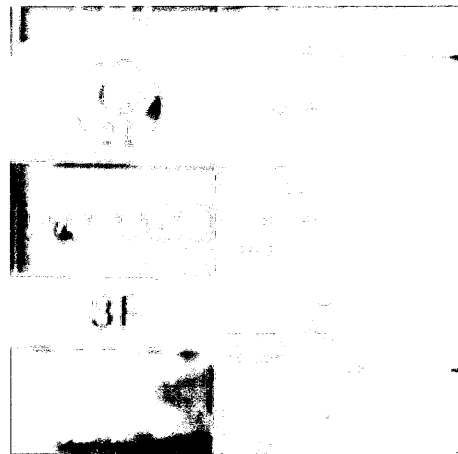


図 4.1: ノイズの影響による色の侵食

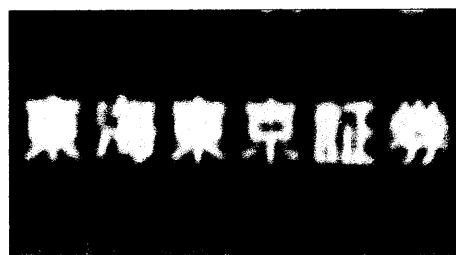


図 4.2: 図(4.1)の部分的な拡大(3倍)

また、膨張は文字領域が小さい場合、途切れている場合には効果があったが、文字と文字の間隔が狭い場合には図(4.5)のように領域を結合してしまい、逆効果となる場合もあり膨張処理の良し悪しは文字の条件によって変わるので情景画像を用いる今回の方法には不向きであると言える。

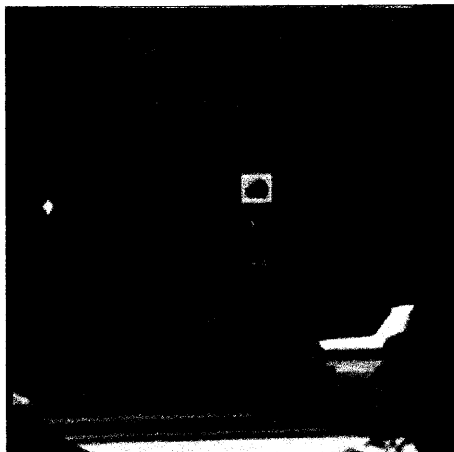


図 4.3: 情景画像



図 4.4: 膨張→縮小未使用



図 4.5: 膨張→縮小使用時の失敗例

縮小の場合も同様で，図 (4.6) のように膨張の場合とは逆に矩形として抽出できていた領域を，領域を削ってしまうことで認識できなくなってしまう場合もあった．



図 4.6: 縮小→膨張使用時の失敗例

全体的に見ても，実験で用いたメディアンフィルタ，膨張・縮小処理自体は良い処理の方法ではあるが，今回の手法に組み込んだ場合にはノイズの影響，領域を大きく変化させるといった影響があり，精度良く抽出するという点では不向きであった．



#### 4.4.2 3枚の切り出し画像から文字列抽出を行う手法

実験の結果、抽出率は全体的に向上していた。これは、前章の図(3.4)から図(3.7)を見ると分かるが、ノイズの影響が1フレームごとに異なるため、3枚の2値画像の領域を統合することでノイズの影響を軽減した2値画像を作成できたと考えられる。また、撮影位置がわずかにずれることによって、3枚の2値画像ではノイズの影響できちんと捉えることのできなかつた領域を重ね合わせて捉えることができたのではないかと考えられる。ただし、これは撮影位置が次のフレームで変化していない場合も考えられるので、必ずしもこの影響が期待できるとは言えない。

## 第5章

### 結論

#### 5.1 まとめ

本論文では動画像からの文字列抽出を、1枚の切り出し画像、3枚の切り出し画像を用いて行う手法を提案した。動画像から画像を1枚切り出して文字列抽出を行う手法では、前章の実験結果の考察で述べたようにあまり良い結果が得られなかった。これは、用いた処理が、今回の抽出の手法には不向きであったからであると言える。動画像から画像を3枚切り出して文字列抽出を行う手法では、松尾らの手法よりも高い抽出率を示した。ノイズの影響があるビデオ映像からの文字列抽出という点では、ノイズの影響が1フレームごとに違うので複数の連続したフレームから画像を切り出してノイズの影響を減少させることが抽出率の向上につながったと言える。

#### 5.2 今後の課題

今回実験を行い、より抽出率を向上させるためなどに必要と思われる今後の課題を以下に挙げる。実験では松尾らの手法を元に文字列の抽出を行ったが、今回は抽出する文字列を一行と限定しているため二行以上文字列が存在する場合にも対応できる方法を考えてみたいと思う。

また、松尾らの論文でも述べられていたことだが、色分解画像が入力時の環境や天候に左右され、安定性に欠けるため、色情報を安定した状態で用いることが可能かどうかの研究も行いたい。さらに、松尾らの手法での細かい点での変更や、他の手法、処理を取り入れた研究を行いたい。

## 謝辞

東北大学大学院工学研究科通信工学科 教授 阿曾弘具先生には、本研究を進めるにあたり多大な御指導を賜りましたことを深く感謝いたします。

東北大学大学院工学研究科通信工学科 助教授 大町真一郎先生には、研究全般にわたって多大な御指導・御助言を賜りましたことを深く感謝いたします。

東北大学大学院工学研究科通信工学科 助手 菅谷至寛氏には、研究に関して御助言を賜りましたことを深く感謝いたします。

平山 勝裕氏には、研究全般に関して広く詳細な御指導・御協力を賜りました。心から感謝致します。

## 参考文献

- [1] 恩田邦夫, 笹尾和仁, 青木由直: “廊下の情景画像からのドアと部屋番号の認識システム,” 信学技報 PRU,88-145 pp.77-84 ,1989 .
- [2] 松尾賢一, 上田勝彦, 梅田三千雄: “カラー情景画像からの文字列領域の抽出,” 電気学会論文誌, Vol.116-C , pp.1252-1258 , 1996
- [3] 藪木 登, 三木 成彦: “道路情景画像中の最高速度標識の検出及び認識,” 信学論, Vol. J77-D, No.7, pp. 1393-1394, 1994.

# ビデオ映像からの 文字列抽出手法に関する研究

東北大学工学部通信工学科  
阿曾研究室 4年 野田口 玄

## 発表の流れ

1. 背景
2. 目的
3. 松尾らの手法
4. 提案手法
5. 実験
6. まとめ
7. 今後の課題

## 1. 背景

- パターン認識の分野で様々な研究が行われ、視覚を用いた認識を行う移動ロボットの実現が可能となってきている
  - 障害物などの比較的単純な物体の認識は可能
  - 文字や記号などの重要な情報を保持しているものの認識は困難
- 文字は、情報の伝達性と保存性に優れており、あらゆる情報を様々な場所で提供する重要なものである



文字を認識することができれば、より高度な移動ロボットの実現が可能

情景中に存在する文字情報を正しく抽出し、理解することが必要

## 1. 背景

- 人間であれば、文字領域を容易に判断し、認識・理解することが可能
- 文字情報を自動で認識する際の問題
  - 文字の方向
  - 文字の大きさ
  - 文字の形状
  - 画像の明るさ
  - 画像の解像度

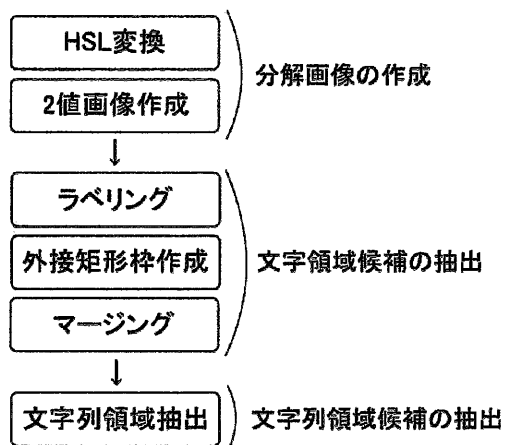
## 1. 背景

- 情景画像からの文字列領域の抽出の従来手法
  - ▣ ドアの部屋番号の文字領域抽出(恩田ら,1989)
  - ▣ 道路情景画像からの速度標識の検出(藪木ら,1994)
    - 背景領域もしくは文字領域が均一色
    - 文字領域の書式が知識として与えられてる
  - ▣ 色彩情報を用いた情景画像からの文字列抽出(松尾ら,1994)
    - デジタルカメラからの静止画像を用いている

## 2. 目的

- 移動ロボットは複雑な情景の動画像の情報を「リアルタイム」で取り入れ認識を行う
- 松尾らの手法を参考にし、デジタルカメラによる静止画像からの文字列抽出とは異なる、抽出の際の問題点を考慮したビデオ映像からの文字列抽出手法を提案
  - ▣ 動画像から画像を1枚切り出し、文字列抽出を行う手法の提案と考察
  - ▣ 動画像から画像を3枚切り出し、文字列抽出を行う手法の提案と考察
  - ▣ ビデオ映像を用いることによる解像度やノイズの影響の考察

### 3. 松尾らの手法



### 3. 松尾らの手法

#### ■ HSL変換

R,G,Bの輝度信号を表す明度 (Lightness)、色の違いを表す色相 (Hue)、色の鮮やかさを表す彩度 (Saturation) の3つの属性に変換

$$L = \frac{R + G + B}{3}$$

$$S = \frac{\text{Max}(R, G, B) - \text{Min}(R, G, B)}{\text{Max}(R, G, B)} \times 255$$

$$H = \frac{G - B}{\text{Max}(R, G, B) - \text{Min}(R, G, B)} \times 60 \quad (\text{Max}(R, G, B) = R)$$

$$H = \left( \frac{B - R}{\text{Max}(R, G, B) - \text{Min}(R, G, B)} + 2 \right) \times 60 \quad (\text{Max}(R, G, B) = G)$$

$$H = \left( \frac{R - G}{\text{Max}(R, G, B) - \text{Min}(R, G, B)} + 4 \right) \times 60 \quad (\text{Max}(R, G, B) = B)$$

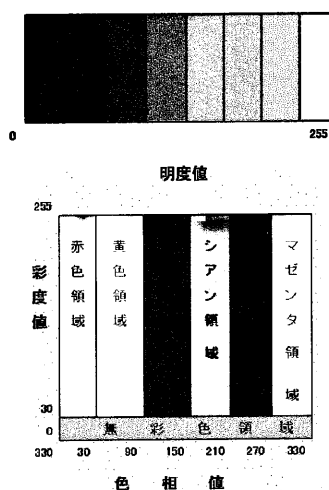


### 3. 松尾らの手法

- カラー情景画像の2値化
  - HSL変換により、情景画像を明度、彩度、色相の3つの特徴に分類
  - 横軸に明度値L、縦軸に横軸の明度値Lに対応した明度値の累積値をとったものを明度ヒストグラム
  - 色相値Hを横軸、彩度値Sを縦軸の尺度にとったものを彩度色相分布図

### 3. 松尾らの手法

- 等間隔分解画像
- ◆ 明度ヒストグラム
  - 等間隔に8分割
  - 明度値Lの1領域の明度範囲は32
  - 分割された領域の範囲の明度値が閉領域となる2値画像が得られる
- ◇ 彩度色相分布図
  - 彩度値30未満を無彩色領域
  - 彩度値30以上では、色相値0度から360度までを、60度刻みで6分割し、2値画像を得る



### 3. 松尾らの手法

- ラベリングによる外接矩形抽出  
各分解画像に対して同一の画素値を持つ閉領域に8連結でのラベリングを行う
- 外接矩形からの文字抽出条件
  - a. 外接矩形の一边が画素数 $p$ 以上かつ縦横の辺比率が1.5未満の矩形は「文字領域候補」
  - b. aにおいて、一边が画素数 $p$ 以上かつ縦横の辺比率が1.5以上10以下の矩形は「文字部分領域候補」
  - c. aまたはb以外の矩形は「不用」とし、対象としないただし、試行実験により、画素数は $p=5$

### 3. 松尾らの手法

- マージング処理  
ある2つの領域の外接矩形の形状と交差状態に基づき、その領域を統合、包括する処理
  1. 重なった矩形同士のみをマージングする交差矩形マージング処理
  2. 分離した矩形同士をマージングする分離矩形マージング処理



- 無条件に交差・分離マージング処理を続けると、文字領域候補の矩形が不用になったり、矩形が立て続けにマージされるなどの問題が起こる

### 3. 松尾らの手法

#### ■ マージング処理

□ 現在の矩形の状態によりマージングの実行を判断する条件を付加

- 2つの外接矩形を  $\alpha$ 、 $\beta$  として、辺比率により3種類の矩形に分類
- どの候補同士が交差しているかを分類し、ランク分けを行う

矩形 $\alpha$	文字領域候補	文字部分領域候補	不 用
矩形 $\beta$	ランクA	ランクB	ランクC
文字領域候補	ランクA	ランクB	ランクC
文字部分領域候補	ランクB	ランクA	ランクC
不 用	ランクC	ランクC	ランクC

- 外接矩形  $\alpha$  と  $\beta$  のマージング後の矩形状態に着目し、可となったものに対してマージングを行う

結合後の矩形状態 矩形のランク	文字領域候補	文字部分領域候補	不 用
ランクA	可	可	不可
ランクB	可	不可	不可
ランクC	不可	不可	不可

### 3. 松尾らの手法

#### ■ マージング処理

$$\bar{L}_\alpha(R) = \frac{\sum L_\alpha(R)}{P_\alpha} \quad \bar{L}_\alpha(G) = \frac{\sum L_\alpha(G)}{P_\alpha} \quad \bar{L}_\alpha(B) = \frac{\sum L_\alpha(B)}{P_\alpha}$$

$$D(\alpha, \beta) = |\bar{L}_\alpha(R) - \bar{L}_\beta(R)| + |\bar{L}_\alpha(G) - \bar{L}_\beta(G)| + |\bar{L}_\alpha(B) - \bar{L}_\beta(B)|$$

矩形  $\alpha$  のラベル上のある画素のRGB値:  $L_\alpha(R), L_\alpha(G), L_\alpha(B)$

ラベルの面積:  $P_\alpha$

ラベル内の平均色:  $\bar{L}_\alpha(R), \bar{L}_\alpha(G), \bar{L}_\alpha(B)$

色差:  $D(\alpha, \beta)$

色差  $D(\alpha, \beta) \leq e$  のものについてマージング処理を行う

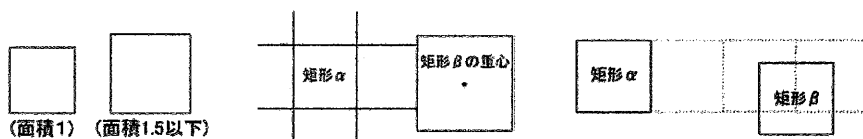
ここでは  $e=30$  とし、文字領域候補と他の領域矩形との併合を防止

### 3. 松尾らの手法

#### ■ 文字列候補領域の抽出

文字の連結性の条件で文字列領域候補の判断を行う

1. 文字領域候補の矩形同士の面積比率が1.5以下である
2. 一つの文字領域候補の矩形が水平または、垂直移動したと仮定し、その移動の軌跡上にもう一つの矩形の重心が存在する
3. 文字領域候補の矩形間の距離が矩形3文字分以下である
4. 文字領域候補の矩形内の対象領域におけるRGB平均値の色差  $D(\alpha, \beta) \leq e$  である



### 3. 松尾らの手法

#### ■ 文字列領域の抽出

各分解画像から抽出された文字列領域候補を全て重ね合わせて文字列統合画像を作成

縦または横に存在する文字列領域を抽出

文字列領域候補の個数を1つとみなし、XまたはY軸に投影

文字列領域候補が交わった総数値を表す文字総数ヒストグラムを得る

### 3. 松尾らの手法

■ 文字総数ヒストグラム

分解画像上で同じ領域に文字列領域候補が得られるほど、文字総数ヒストグラム値が高くなる

$$M_c = \frac{\sum_{x=0}^{m-1} H_x}{m - H_0} \quad H'_x = \begin{cases} H_x - M_c & H_x \geq M_c \\ 0 & H_x < M_c \end{cases}$$

m: 画像サイズの一辺

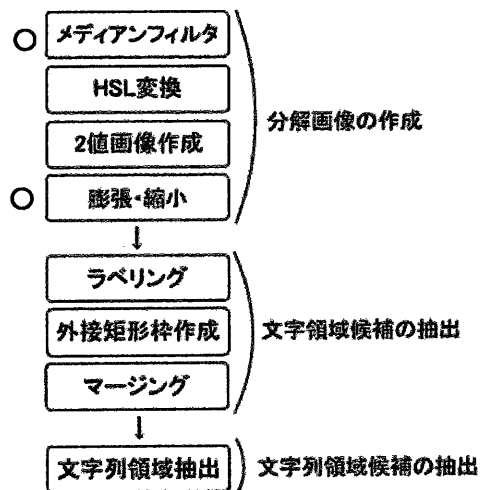
H: X、Y座標の文字総数ヒストグラム値

M<sub>c</sub>: 平均文字数

文字総数ヒストグラムを元に文字列領域候補の抽出を行う

### 4. 提案手法

■ 動画画像から画像を1枚切り出して文字列抽出を行う手法



## 4. 提案手法

- **メディアンフィルタ**  
中心の画素とその周囲8画素の画素の濃度の中央値を中心の画素の値にする  
ノイズを除去し、画像の平滑化に効果がある

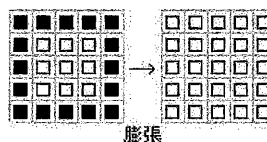
189	170	181
175	181	183
183	178	185

中心の画素の濃度は94,170,175,178,181,183,183,185,189の中央値の181となる

- 分解画像作成の際に2値画像の領域の補正を行うことができる
- 試行実験により、フィルタを5回使用するものとする

## 4. 提案手法

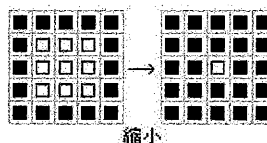
- **膨張・縮小**
  - **膨張**  
有色画素を周囲8近傍に膨張させ、途切れた線などを繋ぐ効果がある
  - **縮小**  
有色画素を周囲8近傍に縮小させ、微小面積の領域を除去する効果がある



- 膨張→縮小、縮小→膨張をそれぞれ1回行う

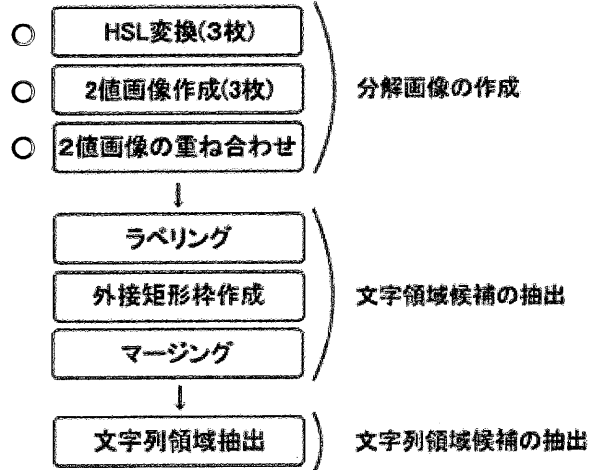
↓

白と黒の微小面積の領域を除去  
途切れた領域を繋ぐ



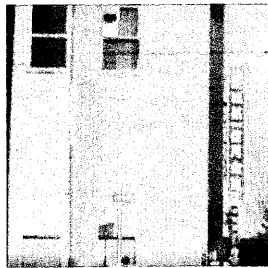
## 4. 提案手法

- 動画像から画像を3枚切り出して文字列抽出を行う手法

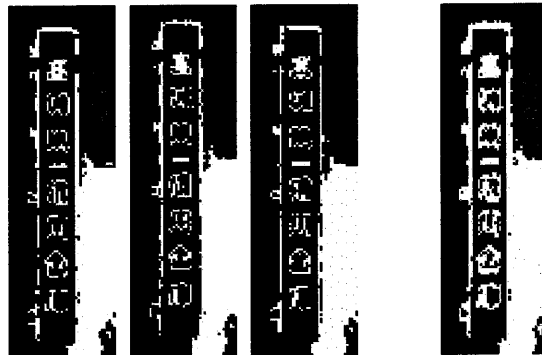


## 4. 提案手法

- 2値画像の重ね合わせ



画像(例)



それぞれの2値画像

重ね合わせ

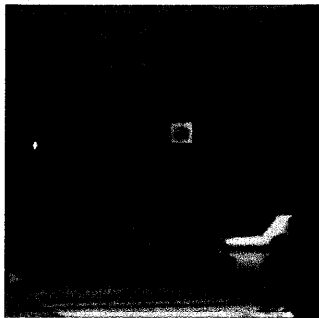
## 5. 実験

- CanonデジタルビデオカメラXV2
- 文字列を含むカラー情景画像20枚(256×256画素、RGB各256階調)
  
- 撮影条件
  - 比較的天候の良い日中
  - 対象文字列には飾り文字や手書き文字が含まれていない
  
- 実験条件
  - 動画像から画像を1枚切り出し文字列抽出を行う手法
    - 松尾らの手法
    - 膨張→縮小処理を組み合わせたもの
    - 膨張→縮小、縮小→膨張処理を組み合わせたもの
    - メディアンフィルタを組み合わせたもの
    - メディアンフィルタ、膨張→縮小処理を組み合わせたもの
    - メディアンフィルタ、膨張→縮小、縮小→膨張処理を組み合わせたもの
  - 動画像から画像を3枚切り出し文字列抽出を行う手法

## 5. 実験

- 文字列抽出の判断は目視により行うものとする

- A: 文字列領域を全て抽出
- B: 文字列領域を他領域も含めて抽出
- C: 文字列領域を部分的に抽出
- D: 文字列領域を他領域も含めて部分的に抽出
- E: 他領域を抽出、または何も抽出できず





## 5. 実験

### ■ 実験結果

手法 抽出率	松尾	膨→縮	膨→縮 縮→膨	メ	メ、膨→縮	メ、膨→縮 縮→膨	3枚画像
A	30%	25%	25%	30%	35%	25%	45%
B	15%	20%	10%	15%	15%	10%	20%
C	25%	35%	35%	35%	20%	20%	15%
D	25%	5%	20%	20%	10%	25%	10%
E	5%	15%	10%	0%	20%	20%	10%
合計	100%	100%	100%	100%	100%	100%	100%

#### 文字列領域の抽出率

- A: 文字列領域を全て抽出  
 B: 文字列領域を他領域も含めて抽出  
 C: 文字列領域を部分的に抽出  
 D: 文字列領域を他領域も含めて部分的に抽出  
 E: 他領域を抽出、または何も抽出できず

## 5. 実験

### ■ 実験結果

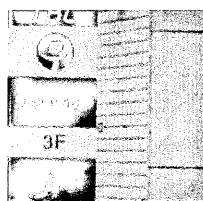
手法 ランクの変化	膨→縮	膨→縮 縮→膨	メ	メ、膨→縮	メ、膨→縮 縮→膨	3枚画像
向上	25%	10%	25%	25%	10%	25%
変化無し	40%	45%	50%	35%	45%	55%
低下	35%	45%	25%	40%	45%	20%
合計	100%	100%	100%	100%	100%	100%

#### 松尾らの手法と提案手法の比較

- ランクが同じ場合でも、抽出した文字数が多い場合には向上  
 抽出した文字数が少ない場合には低下とする

## 6.まとめ

- 動画からの文字列抽出を、1枚の切り出し画像、3枚の切り出し画像を用いて行う手法を提案
- 1枚の切り出し画像から文字列抽出を行う手法
  - ▣ 抽出率は全体的に変化が無いが低下
    - ノイズによる影響で、メディアンフィルタによる平滑化は逆効果
    - 膨張は文字と文字の間隔が狭い場合には領域を結合してしまい逆効果
    - 縮小は小さい文字に対して使用すると候補から除外されてしまう



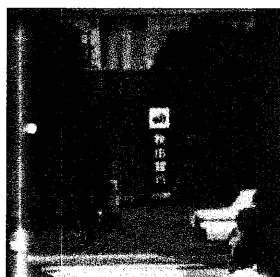
画像(例)



拡大(3倍)

## 6.まとめ

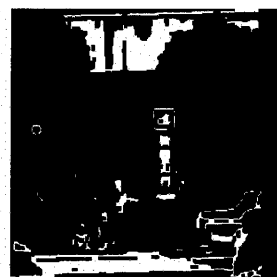
- 動画からの文字列抽出を、1枚の切り出し画像、3枚の切り出し画像を用いて行う手法を提案
- 1枚の切り出し画像から文字列抽出を行う手法
  - ▣ 抽出率は全体的に変化が無いが低下
    - ノイズによる影響で、メディアンフィルタによる平滑化は逆効果
    - 膨張は文字と文字の間隔が狭い場合には領域を結合してしまい逆効果



画像(例)



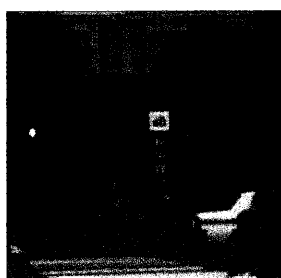
膨張未使用



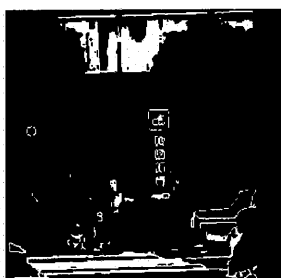
膨張→縮小使用

## 6.まとめ

- 動画像からの文字列抽出を、1枚の切り出し画像、3枚の切り出し画像を用いて行う手法を提案
- 1枚の切り出し画像から文字列抽出を行う手法
  - ▣ 抽出率は全体的に変化が無いが低下
    - ▣ 膨張は文字と文字の間隔が狭い場合には領域を結合してしまい逆効果
    - ▣ 縮小は小さい文字に対して使用すると候補から除外されてしまう



画像(例)



縮小未使用



縮小→膨張使用

## 6.まとめ

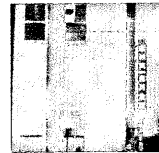
- 3枚の切り出し画像から文字列抽出を行う手法
  - ▣ 抽出率は全体的に向上
    - ▣ ノイズの影響が1フレームごとに異なるため、3枚の2値画像の領域を統合することでノイズの影響を軽減した2値画像を作成できた
    - ▣ 撮影位置がわずかにずれることによって、3枚の2値画像ではノイズの影響できちんと捉えることができなかった領域を重ね合わせて捉えることができる



1フレーム目



2フレーム目



3フレーム目



重ね合わせ



## 7. 今後の課題

- 2列以上文字列がある場合の抽出方法の研究
- 色分解画像が入力時の環境や天候に左右され、安定性に欠けるため、色情報を安定した状態で用いる方法の研究
- 様々な方法を考慮した文字列領域抽出方法の研究